

University of Dundee

MASTER OF DENTAL SCIENCE

Evaluation of the use of the modified Huddart Bodenham & EUROCRAN yardstick for the assessment of surgical outcome for unilateral cleft lip and palate

Patel, Dipali

Award date:
2011

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Evaluation of the use of the modified Huddart Bodenham & EUROCRAN yardstick for the assessment of surgical outcome for unilateral cleft lip and palate

Dipali Patel

2011

University of Dundee

Conditions for Use and Duplication

Copyright of this work belongs to the author unless otherwise identified in the body of the thesis. It is permitted to use and duplicate this work only for personal and non-commercial research, study or criticism/review. You must obtain prior written consent from the author for any other use. Any quotation from this thesis must be acknowledged using the normal academic conventions. It is not permitted to supply the whole or part of this thesis to any other person or to post the same on any website or other online location without the prior written consent of the author. Contact the Discovery team (discovery@dundee.ac.uk) with any queries about the use or acknowledgement of this work.

**EVALUATION OF THE USE OF THE MODIFIED HUDDART BODENHAM &
EUROCRAN YARDSTICK FOR THE ASSESSMENT OF SURGICAL
OUTCOME FOR
UNILATERAL CLEFT LIP AND PALATE**



**A DISSERTATION PRESENTED FOR THE DEGREE OF
MASTER OF DENTAL SCIENCE**

DIPALI PATEL

BDS

MFDS RCSGlas

MOrth RCSEd

**UNIVERSITY OF DUNDEE
JANUARY 2011**

TABLE OF CONTENTS – OUTLINE

TITLE	i
TABLE OF CONTENTS – OUTLINE.....	ii
TABLE OF CONTENTS – DETAIL	iii
LIST OF TABLES.....	v
LIST OF FIGURES	vi
LIST OF APPENDICES.....	vii
ACKNOWLEDGEMENTS	viii
DECLARATION.....	ix
CERTIFICATE	x
ABSTRACT	xi

CHAPTER ONE:	INTRODUCTION
CHAPTER TWO:	LITERATURE REVIEW
CHAPTER THREE:	MATERIALS AND METHOD
CHAPTER FOUR:	AIMS, OBJECTIVES AND HYPOTHESIS
CHAPTER FIVE:	RESULTS
CHAPTER SIX:	DISCUSSION
CHAPTER SEVEN:	CONCLUSION
CHAPTER EIGHT:	REFERENCES
CHAPTER NINE:	APPENDICES & INDEX

TABLE OF CONTENTS - DETAIL

CHAPTER ONE: INTRODUCTION	1
CHAPTER TWO: LITERATURE REVIEW	4
2.1 Embryology of cleft lip and palate.....	4
2.1.1 Normal development of the lip and palate	4
2.1.2 Pathogenesis of the formation of primary and secondary palate	7
2.1.3 Aetiology of cleft lip and palate.....	9
2.2 Classification of cleft lip and palate	11
2.2.1 Veau 1931	11
2.2.2 Kernahan and Stark 1958.....	11
2.2.3 Kriens 1989	13
2.3 Impact of orofacial clefts on individuals.....	16
2.4 Current management of UCLP	19
2.5 Standards of care.....	28
2.5.1 A Six Centre European Comparison Study : Eurocleft Study.	28
2.5.2 Eurocleft project 1996-2000	29
2.5.3 Clinical Standards Advisory Group (CSAG, 1998).....	30
2.5.4 The importance of measuring surgical outcome.....	31
2.6 Indicators of & methods for assessing surgical outcome	33
2.6.1 Dental arch relationship	33
2.6.2 Indices	34
2.6.3 Antero-Posterior Skeletal Bases	35
2.6.4 Lateral Cephalometry.....	35
2.6.5 Soft tissue profile	36
2.6.6 Photographs.....	36
2.6.7 Stereophotogrammetry.....	37
2.6.8 Cephalometric radiography.....	38
2.7 The choice of study models for outcome assessment.....	40
2.8 Different types of indices	41
2.8.1 GOSLON	41
2.8.2 5-year-old index	45
2.8.3 Huddart Bodenham (1972)	46
2.8.4 The Modified Huddart Bodenham Index	48
2.8.5 EUROCRAN Yardstick (Oskouei, 2007)	50
2.9 The choice of 5-year-olds.....	52
Summary	55

CHAPTER THREE: AIMS, OBJECTIVES & HYPOTHESIS	56
CHAPTER FOUR: MATERIALS AND METHOD	57
4.1 Materials	57
4.2 Examiners.....	57
4.3 Method	59
4.3.1 Modified Huddart Bodenham (see appendix I to IV)	60
4.3.2. EUROCRAN Yardstick (see appendix VI to XI)	61
4.3.3 Examiners	62
4.3.4 Timing.....	63
4.3.5 Likert Scale	63
4.4 Statistical Analysis	64
CHAPTER FIVE: RESULTS.....	68
5.1 MODIFIED HUDDART BODENHAM.....	69
5.2 EUROCRAN Yardstick	80
5.3 Learning and Scoring	82
5.4 User Friendliness.....	85
5.5 Results of the pilot study to test the supplemental flowchart	86
CHAPTER SIX: DISCUSSION	88
6.1 Reliability	88
6.1.1 Intra-examiner reliability	88
6.1.2 Inter-examiner reliability	89
6.1.3 Calibration.....	89
6.1.4 Professionals Complimentary to Dentistry (PCDs)	90
6.1.5 Likert Scale	90
6.1.6 Timing.....	93
6.1.7 Biases and criticism of the study design	93
6.1.8 Summary	95
CHAPTER SEVEN: CONCLUSION	98
RECOMMENDATIONS FOR FUTURE RESEARCH	99
CHAPTER EIGHT: REFERENCES	100

LIST OF TABLES

Table 1	Further description of cleft lip and palate
Table 2	Unweighted kappa statistics
Table 3	Intra-examiner reliability for modified Huddart Bodenham assessed using intraclass correlation coefficient
Table 4	Statistical significance for intra-examiner reliability for modified Huddart Bodenham
Table 5	Mean difference, standard deviation and limits of agreement
Table 6	Inter-examiner reliability for modified Huddart Bodenham assessed using intraclass correlation coefficient
Table 7	Statistical significance for inter-examiner reliability for modified Huddart Bodenham
Table 8	Intra-examiner reliability of antero-posterior dimensions for EUROCRAN Yardstick
Table 9	Intra-examiner reliability of the palatal vault for EUROCRAN Yardstick
Table 10	Inter-examiner reliability of antero-posterior dimensions for EUROCRAN Yardstick
Table 11	Inter-examiner reliability of the palatal vault for EUROCRAN Yardstick
Table 12	Time taken to learn to apply each index
Table 13	Time taken to score each index
Table 14	Total time taken for the entire exercise
Table 15	Likert scale
Table 16	Inter-examiner reliability for the pilot study
Table 17	Data for the pilot study, using mHB to test the supplemental flowchart
Table A	To show statistical significance at 5% level for inter-examiner reliability for the pilot study

LIST OF FIGURES

- Figure 1** Frontal view of a 4.5 week embryo
- Figure 2** Craniofacial region of the developing embryo during the
A. fifth week B. sixth week in utero.
- Figure 3** A 7-week embryo. Maxillary prominences have fused with the medial nasal prominences. B. 10-week embryo.
- Figure 4** Frontal section through the head of 6-week-old embryo.
- Figure 5** Elevation of the lateral palatal shelves in a 7.5- week embryo.
- Figure 6** Fusion of the palatine shelves and the nasal septum in a 10-week embryo.
- Figure 7** Kernahan's striped Y classification
- Figure 8** Cleft lip and palate. (a) cleft lip and alveolus (b) cleft palate (c) incomplete unilateral cleft lip and palate (d) complete unilateral cleft lip and palate (e) complete bilateral cleft lip and palate
- Figure 9** Pictorial chart for modified Huddart Bodenham (mHB)
- Figure 10** Standardised preparation of study models, according to the Eurocleft recommendations
- Figure 11** Boxplots to show normal distribution for first mHB
- Figure 12** Boxplots to show normal distribution for first mHB
- Figure 13** Bland & Altman graph for examiner A's intra-examiner reliability
- Figure 14** Bland & Altman graph for examiner B's intra-examiner reliability
- Figure 15** Bland & Altman graph for examiner C's intra-examiner reliability
- Figure 16** Bland & Altman graph for examiner D's intra-examiner reliability
- Figure 17** Bland & Altman graph for examiner E's intra-examiner reliability
- Figure 18** Bland & Altman graph for examiner F's intra-examiner reliability
- Figure 19** Boxplots to show normal distribution of data for dentists & PCDs for mHB
- Figure 20** Boxplots to show normal distribution of data for calibrated & non-calibrated examiners for mHB
- Figure 21** Boxplots to show asymmetric distribution of data for all four examiners
- Figure A** Boxplot to show asymmetric distribution of data from table 12
- Figure B** Boxplot to show asymmetric distribution of data from table 13

LIST OF APPENDICES

- I** Instruction sheet for scoring study models using the modified Huddart Bodenham
- II** Exceptions to the rule and scoring form for modified Huddart Bodenham
- III** Pictorial flowchart for modified Huddart Bodenham
- IV** Written flowchart for modified Huddart Bodenham
- V** Final checklist before data is handed in
- VI** Instructions and score sheet for scoring study models using EUROCRAN Yardstick
- VII** Definitions of Eurocran Primary Dentition Yardstick
- VIII** Written flowchart for EUROCRAN Yardstick: Anteroposterior/vertical/transverse scoring
- IX** Examples of grades 1 to 4 using photos of study models for Anteroposterior/vertical/transverse aspects
- X** Written flowchart for EUROCRAN Yardstick: Palatal form scoring
- XI** Examples of grades 1 to 4 using photos of study models for palatal aspects
- XII** Final checklist before data is handed in
- XIII** Is this method of scoring models user friendly
- XIV** Raw data for modified Huddart Bodenham
- XV** Raw data for EUROCRAN Yardstick for calibrated examiners
- XVI** Raw data for EUROCRAN Yardstick for non- calibrated examiners
- XVII** Raw data for the pilot study to test the written flowchart for modified Huddart Bodenham

INDEX

- mHB modified Huddart Bodenham Index
- EY EUROCRAN Yardstick

ACKNOWLEDGEMENTS

Firstly, I would like to thank Gary Grime, David M^cMahon, Leslie Milne, Professor Ricketts and Kevin Wilson, I am indebted to them for all the time they spent scoring the study models. I would also like to thank Professor Mossey for all the support, time and enthusiasm he has given me during the course of this thesis. Thanks also to Mairi Milne at CleftSiS for her help with the study models. I also appreciate the statistical advice given by Professor Donnan. Finally, I would like to thank my mum, I could not have completed this thesis without her endless patience and support.

DECLARATION

I, Dipali Patel, declare that the following dissertation
is entirely my own work

Signed:

Date:

CERTIFICATE

I hereby certify that this candidate has fulfilled the condition of Ordinance Number 12 and the Regulations of the University of Dundee for the Degree of Master of Dental Science.

Signed by:

Date:

ABSTRACT

Background: The measurement of maxillary arch constriction in patients born with cleft lip and /or palate for assessment of treatment outcome is a well accepted concept. In order to improve surgical outcome, it is essential to have a valid and reliable tool to assess results. Ongoing, international, multi-centre, randomised clinical trials are likely to produce subtle, but significant results between different surgical techniques, timings and surgeons. Therefore, it is essential to compare emerging indices to assess which best fits the requirements set by the World Health Organisation.

Objectives: The aim of this study was to compare the reproducibility of two indices: modified Huddart Bodenham (mHB) and EUROCRAN Yardstick .

Design: Retrospective, case-control study of study models of 5-year-old patients with unilateral cleft lip and palate.

Method : Thirty study models were scored using mHB and EUROCRAN Yardstick by six examiners on two occasions at least one month apart.

Results: The intra- and inter-examiner intraclass correlation reproducibility with mHB indicated good reproducibility (range 0.75 to 0.94). The Bland & Altman graphs confirmed this and did not show any areas of systematic bias. The intra- and inter-examiner reproducibility using kappa statistics for EUROCRAN yardstick indicated wide variability (range 0.33 to 0.95) for maxillary arch constriction. The intra- and inter-examiner reproducibility using kappa statistics for EUROCRAN Yardstick indicated poor reproducibility (range -0/06 to 0.50) for palatal scores. The best score of 0.94 with mHB was 0.80 at the second reading and the best score of 0.95 with EUROCRAN Yardstick was 0.41 at the second reading.

Conclusion: This study has revealed that mHB is more reliable than EUROCRAN Yardstick. It is superior in terms of user friendliness and time taken to learn to use the

index. However, the study showed that it takes longer to score the study models using mHB than EUROCARN Yardstick, but this improves at second exposure.

CHAPTER ONE: INTRODUCTION

Orofacial clefts are the commonest craniofacial anomaly. (Gorlin et al., 1971). These clefts involve the lip and/or palate (CLP) or isolated clefts of the palate (CP). On a worldwide level orofacial clefts affect approximately 1 in every 600 newborn babies (Mossey and Little, 2002c) with significant ethnic variability, with Asians being affected most commonly and the least common in African people. A child born with orofacial clefting will require complex long-term surgical and non-surgical treatment, depending upon the severity of the cleft. There may be lifelong implications for these individuals, resulting in considerable disruption to their lives and often, adverse psychological consequences to themselves and their families.

Of the various types of orofacial clefts (see chapter 3), cleft lip and palate is the most complex to treat and has the greatest level of morbidity associated with it. The location of the cleft lip and palate anomaly has an impact on speech, external facial appearance, dental relationship, craniofacial growth, hearing, and via these handicaps, to possible social or psychological impairment. Inevitably, bilateral cleft lip and palate (BLCP), is more severe in terms of morbidity and complexity of treatment. The impact and disruption it causes to these patients is far more serious than unilateral cleft lip and palate (UCLP), but fortunately BCLP is rarer than UCLP. A study compared the soft tissue parameters of non-clefts patients with patients with UCLP and UCL using stereophotogrammetry. It showed that the soft tissue parameters of patients with UCL was more like the non-cleft group but significantly different to patients with UCLP (Hood et al., 2004), thus confirming the effect of UCLP on facial appearance.

The principal objectives of treatment for individuals with UCLP are to improve feeding, hearing, speech, facial appearance and overall to reduce morbidity and the negative psychological impact to these individuals. If these objectives are achieved, they maximise the chances of a child with UCLP growing up and developing normally

within their social environment. The early stages of treatment for individuals with UCLP involves primary surgical repair of the lip and palate. There are various surgical techniques and to date there is not an agreed optimal surgical technique (Molsted et al., 1992) or timing for cleft repair (Rohrich et al., 1996). Poorly performed surgery carries a high risk of compromising facial growth, dental development and speech impairment (Roberts et al., 1991).

The World Health Organisation (WHO) and the American Cleft Palate Association (ACPA) have both published national standards on management of oral clefts in terms of surgical repair and non-surgical treatment. Surgical improvement is a necessity in the field of evidence-based medicine. Clinical governance also demands that standards of treatment should be monitored. In order to monitor standards and assess improvement, it is essential to have a reliable measurement tool for the assessment of surgical outcome.

A European, multicentre clinical audit of treatment outcome for individuals with complete UCLP demonstrated that it is possible to detect differences in outcome using the GOSLON Yardstick to assess dental arch relationships. (Shaw et al., 1992b) and hence, surgical outcome.

A number of methods have emerged, which measure surgical outcome by measuring maxillary arch constriction. These are: the GOSLON yardstick (Mars et al., 1987), the 5-year-old index (Atack et al., 1997a), the modified Huddart/Bodenham (Mossey et al., 2003, Gray and Mossey, 2005) and the EUROCRAN Yardstick (Oskouei, 2007).

The GOSLON yardstick has been used widely to date but there is a perceived need for improved sensitivity and objectivity. The GOSLON yardstick uses a categorical scoring system designed to assess patients aged 10 with UCLP. The advantage of the 5-year-old index is that it measures the outcome based on surgical intervention alone, lending itself to a more accurate basis for surgeons to detect differences in

surgical outcome. Whereas, by age 10, any further treatment such as alveolar bone grafting or orthodontics, dilutes the effects of surgery alone. However, the 5-year-old index also uses a categorical scoring system as it is based on the GOSLON yardstick. Both methods require a calibration course and recalibration is necessary to ensure consistency. Ten reference models for each index have to be used for comparison during the scoring of the study models, all of which adds to the complexity of the exercise.

Future and ongoing surgical trials are likely to produce subtle, but perhaps significant results between different techniques, timing and surgeons. Therefore there is a need for a discriminating measure of surgical outcome to ensure that small differences can be detected. These small differences are more likely to be detected from using a continuous scale rather than a categorical scale. Both, the modified Huddart Bodenham and the Eurocran Yardstick use a continuous scale. Their main difference is that the latter evaluates the arch in anterior-posterior (A-P), vertical, transverse and palatal morphology where as the former evaluates the A-P and transverse planes, alone.

These two indices have emerged following criticism and perceived shortcomings of the GOSLON Yardstick and the 5-year old index. Therefore there is now a need to compare the two emerging indices.

CHAPTER TWO: LITERATURE REVIEW

2.1 Embryology of cleft lip and palate

Orofacial clefts occur when tissues that form the lip and palate fail to fuse during normal development of the embryo. There are two types of cleft anomaly; cleft lip with or without palate and isolated palate. Both result from fusion at two different stages of dentofacial development.

Cleft lip is a result of failure of fusion (*at 4-6 weeks in utero*) between the medial nasal, lateral nasal and maxillary swellings.

Cleft palate is a result of failure of fusion (*at 8 weeks in utero*) of the lateral palatal swellings.

2.1.1 Normal development of the lip and palate

In order to understand the pathogenesis of orofacial clefting, it is useful to understand the normal embryological processes involved in the formation of the lip and palate.

Formation of the primary palate

At the end of the fourth week, facial prominences consisting primarily of neural crest-derived mesenchyme are formed from the first pharyngeal arch. These are the maxillary prominences (lateral to the stomodeum), mandibular prominences (caudal to the stomodeum) and the frontonasal prominence (upper border of stomodeum). On both sides of the frontonasal prominences, local thickenings of surface ectoderm, nasal placodes, originate (Figure 1).

The embryonic period begins from 3rd to 8th weeks of development and is the time when each germ layer, ectoderm, mesoderm and endoderm, gives rise to a number of specific organs and tissues.

Neural crest cells are important for craniofacial development because they contribute to many structures in this region. They originate in the neuroectoderm to form the facial skeleton and most of the skull. These cells constitute a vulnerable population

as they leave the neuroectoderm; they are often the target for teratogens. Therefore it is not surprising that craniofacial abnormalities are common birth defects.

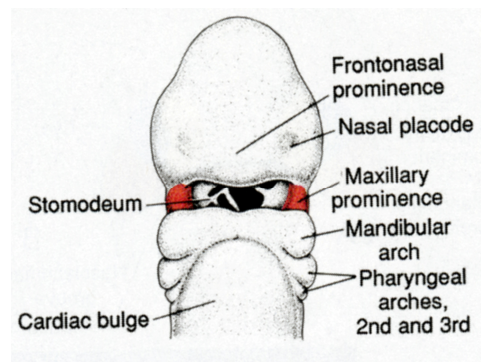


Figure 1: Frontal view of a 4.5 week embryo, adapted from (Sadler, 2004)

During the fifth week, the nasal placodes invaginate to form nasal pits. Thus, they create a ridge of mesenchymal tissue at its periphery, which proliferates to form nasal prominences, the medial and lateral, the former is on the inner edge of the pits (Figure 2).

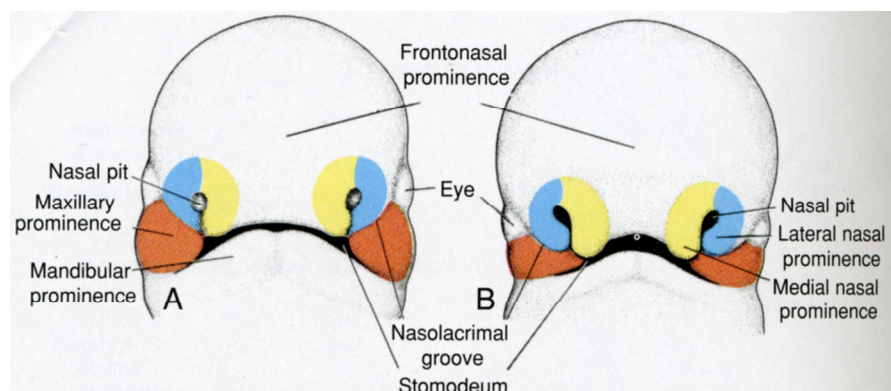


Figure 2: Craniofacial region of the developing embryo during the A. fifth week B. sixth week in utero, adapted from (Sadler, 2004)

Over the next two weeks, the maxillary prominences continue to increase in size. Simultaneously, they grow medially, compressing the medial nasal prominences toward the midline. Subsequently the cleft between the medial nasal prominence and the maxillary prominence is lost, and the two fuse (Figure 3). Hence, the upper lip is formed by the two medial nasal prominences and the two maxillary prominences. The lateral nasal prominences do not participate in formation of the upper lip, as they go on to form the alae of the nose.

A partial or complete lack of fusion of the maxillary prominence with the medial nasal prominence may lead to clefting of the lip and alveolus. This may be unilateral or bilateral.

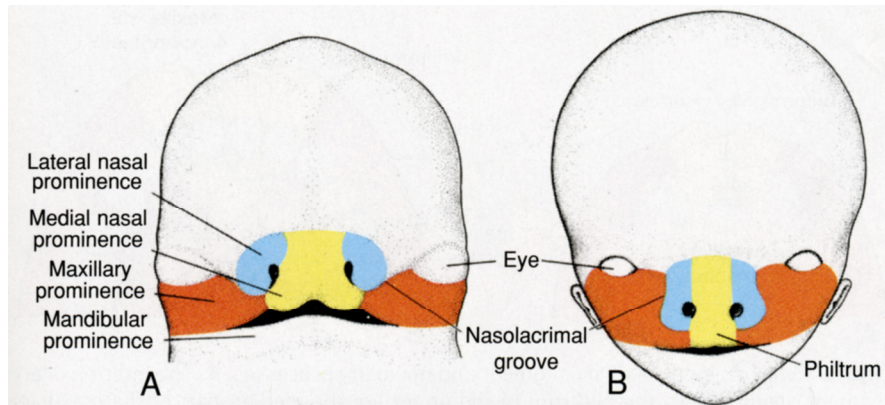


Figure 3: A 7-week embryo. Maxillary prominences have fused with the medial nasal prominences. B. 10-week embryo, adapted from (Sadler, 2004)

Formation of the Secondary Palate

Formation of the secondary palate begins early in the sixth week of gestation with the development of two mesenchymal projections extending from the internal aspect of the maxillary prominences. These are called the lateral palatine shelves and initially are directed obliquely downward on each side of the tongue (Figure 4).

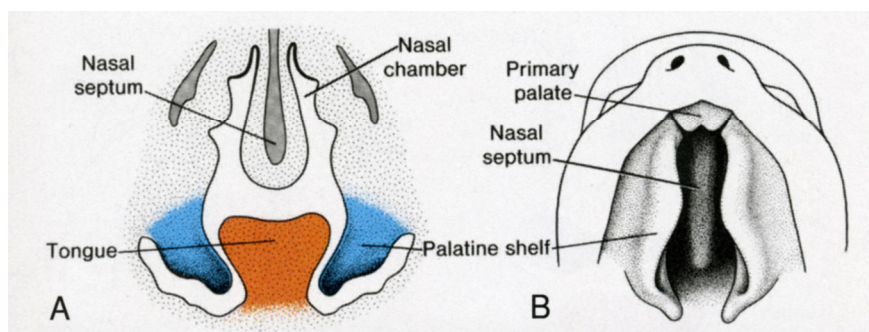


Figure 4: Frontal section through the head of 6-week-old embryo. A. the palatine shelves are in the vertical position on each side of the tongue. B. note the clefts between the primary triangular palate and the palatine shelves, which at this stage are still vertical, (adapted from Sadler 2004)

At seven to eight weeks, the lateral palatine shelves elongate and elevate to a horizontal position above the tongue. The tongue meanwhile becomes relatively small and moves inferiorly (Figure 5).

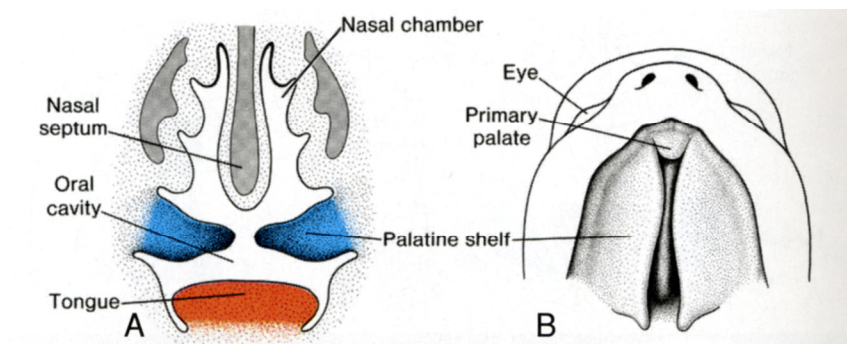


Figure 5: Elevation of the lateral palatal shelves in a 7.5-week embryo, adapted from (Sadler, 2004)

Once, elevated, the lateral palatine shelves approach one another and fuse by combination of cell death and cell migration. They also fuse with the nasal septum and the posterior part of the primary palate (Figure 6).

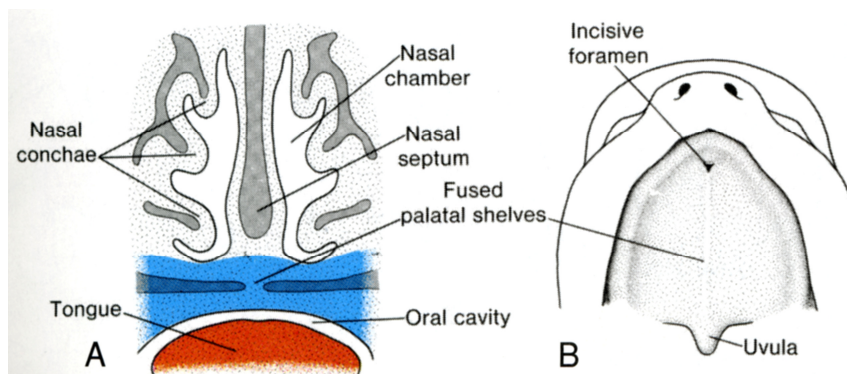


Figure 6: Fusion of the palatine shelves and the nasal septum in a 10-week embryo, adapted from (Sadler, 2004)

2.1.2 Pathogenesis of the formation of primary and secondary palate

The facial processes contain mesenchyme bounded by epithelium, and in order for union to occur, this epithelial barrier must disappear. The mechanism of union

between processes is thought to either involve merging or transformation of the barrier of epithelium into mesenchyme, allowing mesenchymal continuity.

Proposed mechanisms for joining of prominences

One mechanism states that when one prominence is in approximation with another, ectodermal resorption at the point of contact and fusion is the result. For example, when maxillary prominences contact the labial component of the medial nasal prominence, ectodermal resorption at the point of contact occurs resulting in union of the prominences (Ferguson, 1987) & (Ferguson, 1988).

Others believe that the prominences are swellings, corresponding to mesenchymal growth centres that lead to merging between prominences, not fusion. The grooves between prominences were believed to flatten out due to proliferation of the underlying mesenchyme (Ferguson, 1987) & (Ferguson, 1988).

The forces generating the elevation of the lateral palatine shelves is thought to be caused by an intrinsic force, which overcomes resisting factors such as the tongue. This force is generated by synthesis of extra-cellular matrix molecules, one of which is hyaluronic acid, which is capable of binding ten times its own weight of water. Swelling of the mesenchymal cells due to hydration of hyaluronic acid therefore causes an elevating force. The mesenchymal cells themselves are also contractile and may proliferate differentially resulting in further elevating force (Ferguson, 1988).

It has been suggested that isolated cleft palate is more common in females than males because transposition of the palatal shelves occurs one week later in the female embryo and so there is a greater opportunity for an environmental insult to affect successful elevation (Burdi and Silvey, 1969) (Ferguson, 1987).

Cleft palate may result from disturbances at any stage of palate development. The disruption may be due to defective palatal shelf growth, delayed or failed shelf elevation, defective shelf fusion, failure of medial edge cell death, postfusion rupture or failure of mesenchymal consolidation differentiation (Ferguson, 1987). Cleft palate may occur in association with clefting of the lip and alveolus, if

there is disruption of formation of both the primary and secondary palate. Clefting of the lip or palate may also each occur in isolation.

2.1.3 Aetiology of cleft lip and palate

For fusion to take place, the breakdown of the overlying epithelium is followed by invasion of mesenchyme. If this process is to take place successfully a number of factors need to interact at the right time. For example, an inherited tendency towards short palatal shelves can be compensated by over- development of other factors. However, if one of these factors is also affected or an environmental insult occurs at the time that palate formation is taking place, a cleft may result. Therefore a cleft lip and palate is described as having a polygenic inheritance with a threshold. Environmental risk factors include maternal alcohol intake, smoking, phenytoin intake, folic acid deficiency and steroid therapy. All of these factors may precipitate a susceptible foetus towards the threshold.

In summary orofacial clefts are multifactorial with both genetic and environmental influences.

Orofacial clefts may occur in isolation or as part of a syndrome. Over 300 syndromes are recognised (involving the oral, cardiac, skeletal and other body areas). Fifteen per cent of all orofacial clefts are syndromic (Mossey and Little, 2002b) It is generally accepted that associated syndromes occur less frequently in infants who have cleft lip and palate than in those with isolated cleft palate. Cleft lip is the least associated with syndromes (Mossey and Little, 2002a). BCLP is almost twice as likely to be associated with syndromes than UCLP (Wolfsberg, 2002). Approximately 50% of syndromes are associated with isolated cleft palate and 25% with cleft lip and palate (Soal, 2002). The most common syndromes associated with isolated cleft palate are, Pierre Robin syndrome, Stickler, Van Der Woude, Aperts, Crouzons and Fetal alcohol syndrome. Cleft palate is more common in patients with Aperts than Crouzons syndrome

Treacher-Collins syndrome has full penetrance and variable expressivity and so isolated cleft palate does not always present in these patients. Isolated cleft palate often presents in patients with Pierre Robin syndrome, as the tongue tends to occupy the palatal area and thus prevents the palatal shelves from merging, resulting in isolated cleft palate. Pierre Robin syndrome, often presents with Stickler syndrome and hence the association with isolated cleft palate and Stickler syndrome. Van Der Woude Syndrome is the only which presents with either isolated cleft palate or unilateral cleft lip and palate, whereas other associated syndromes will only present with a specific cleft type.

2.2 Classification of cleft lip and palate

Oral clefts can vary widely, from a simple cleft lip to bilateral cleft of the lip and the entire palate. In some way, each patient with a cleft of the lip and/or palate is different from every other. However, it is important to classify patients into groups for the purposes of management and research studies. Many different methods of classifying cleft lip and palate have been suggested by (Veau, 1931), (Kernahan, 1958) and (Kriens, 1989).

2.2.1 Veau 1931

Veau proposed a method of classification that categorised clefts into four categories.

- Clefts of the soft palate only
- Clefts of the soft and hard palate
- Complete unilateral clefts of the lip and palate
- Complete bilateral clefts of the lip and palate

This classification is simple and is still in use in some units today. However, it does not take into account clefts of the lip alone or incomplete clefts.

2.2.2 Kernahan and Stark 1958

This classification is based on embryology that uses the incisive foramen as a boundary, dividing clefts of the primary palate from those of the secondary palate. The primary palate refers to the lip, alveolus and the palate anterior to the incisive foramen. A complete cleft of the primary palate will involve the full thickness of these structures. The secondary palate refers to the soft and hard palate, up to the incisive foramen.

This idea later became a symbolic method, in the development of Kernahan's striped Y symbolic classification (Kernahan, 1971) (Fig 7) This classification was developed

to overcome many of the disadvantages of verbal or numerical classifications, and to allow identification at a glance of the true preoperative condition of the patient.

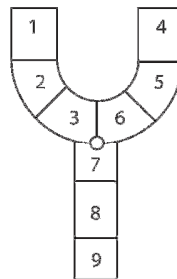


Figure 7: Kernahan's striped Y classification

Areas 1 and 4 represent the right and left sides of the lip, respectively. The alveolus is represented by areas 2 and 5. The hard palate anterior to the incisive foramen by areas 3 and 6. The hard palate posterior to the incisive foramen by areas 7 and 8. and the soft palate by area 9.

The clefts are further described according to the depth of structures affected and the number of sides. (Table 1 & Figure 8).

Clefts of the primary palate	Clefts of the secondary palate	Clefts of the primary and secondary palate
Unilateral (left or right)	Complete	Unilateral (left or right)
Complete	Incomplete	Complete
Incomplete	Submucous	Incomplete
Bilateral		Bilateral
Complete		Complete
Incomplete		Incomplete

Table 1: Further description of cleft lip and palate

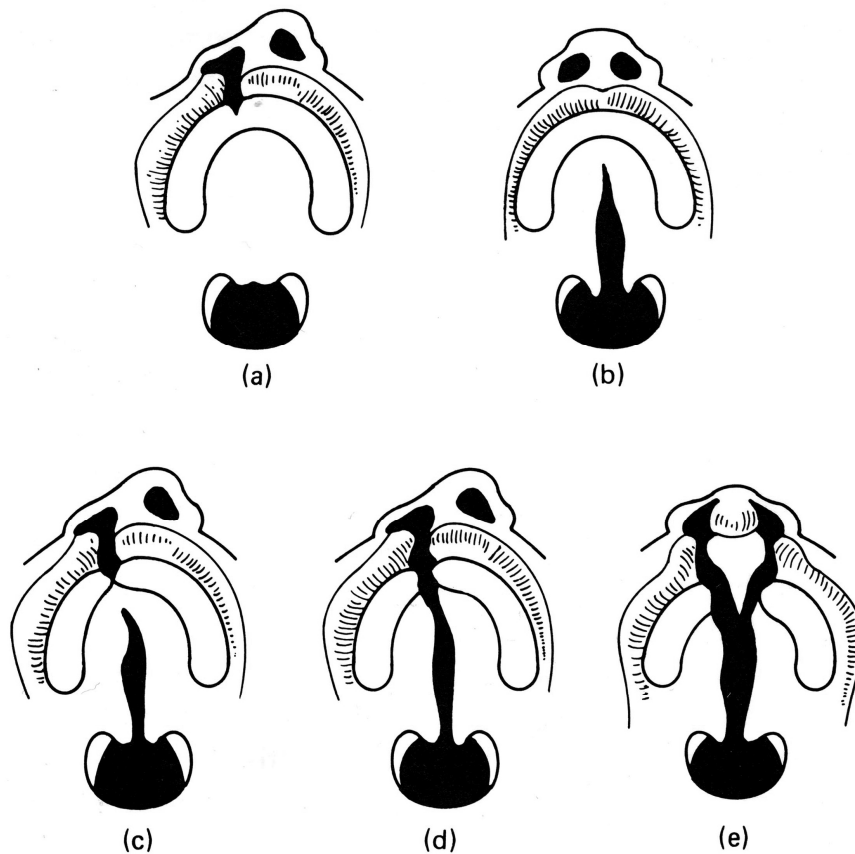


Figure 8: Cleft lip and palate. (a) cleft lip and alveolus (b) cleft palate (c) incomplete unilateral cleft lip and palate (d) complete unilateral cleft lip and palate (e) complete bilateral cleft lip and palate

2.2.3 Kriens 1989

In the UK, there has been a general move to adopt a simple system, which is easy to use yet has sufficient accuracy for most purposes. It is based on the letters of the palindrome LAHSHAL, which represent the two sides of the lip (L), alveolus (A), hard palate (H) and the soft palate (S). Upper and lower case letters denote complete and incomplete clefts, respectively. It has been modified on the recommendation of the Royal College of Surgeons by omitting one “H” (this removes the scope to record bilateral clefts of the hard palate, but results in a far simpler system). It is compatible with the WHO classification of diseases (ICD 10) and allows clefts to be coded for computer use. This is pertinent for cleft epidemiology and for registers of cleft births including the Craniofacial Anomalies Register (CRANE), which has the aim of

recording every patient with cleft in England and Wales and the CLEFTSIS Electronic Patient Record (EPR) in Scotland.

Submucous clefts of the palate probably represent inadequacy of mesenchymal ingrowth following epithelial fusion of the palate and classically present with a bifid uvula, notch at the back of the hard palate and a lucent line along the palate midline with misalignment of the palatal muscles.

Prevalence

On a worldwide level, orofacial clefts affects 1 in 600 newborn babies and the prevalence of cleft lip with or without palate is much higher than isolated cleft palate (Mossey and Little, 2002a). UCLP is the most frequent single type of cleft, accounting for about 30-35% of cases. Cleft lip and cleft palate, each account for between 20-25% and BCLP is the most rare at 10% with submucous and other clefts accounting for the rest (Mossey and Little, 2002b). It is difficult to report an accurate figure of the prevalence for orofacial clefts and its different subtypes because:-

1. In many parts of the world, e.g. Africa and Asia, this data is difficult to obtain as birth registration is not imperative.
2. Like for like comparison is not feasible because of the lack of and/or failure to apply an internationally comparable orofacial classification.
3. There is a great deal of geographical and ethnic variation, where it is most common amongst children who are of North European and Asian origin and less common in African origin. Both of these features are more apparent in patients with cleft lip and palate than those with isolated cleft palate.

Cleft lip and palate

A family history can be found in around 40% of cases of cleft lip with or without palate and the risk of unaffected parents having another child with this anomaly is 1 in 20. Males are affected more frequently than females and the left side is involved more commonly than the right (Mitchell, 2007).

Isolated cleft palate

This cleft type has a lesser genetic component with a family history in around 20% and a reduced risk of further affected offspring to normal parents (1 in 80). In contrast to cleft lip and palate it affects females more than males (Mitchell, 2007).

2.3 Impact of orofacial clefts on individuals

Clinical Problems in Patients with Orofacial Clefts

Feeding is difficult due to communication between the oral and nasal cavities. A retrospective investigation in the UK reported that 25% of patients with non-syndromic orofacial clefts had significant feeding problems. Their mean weight gain was considerably lower than expected for an infant free from cleft (Jones, 1988). Another prospective study also found that poor feeding skills were detected in one third of newborns, but that the prevalence of poor feeding reduced with age (Reid et al., 2006). Both studies infer that the best feeders were patients with cleft lip only and worst were those in the cleft palate and cleft lip and palate groups.

Hearing problems are secondary to middle ear drainage due to Eustachian tube dysfunction. This results from a cleft involving the posterior part of the hard and soft palate. The tensor palatine muscles of the soft palate are connected to the Eustachian tube. The lack of muscle activity in these muscles secondary to the cleft leads to a lack of drainage of the middle ear, which results in middle ear infections and eventually bursts the ear drum. The majority of these patients therefore undergo placement of grommets in the ear drum to enhance drainage of the middle ear. Hearing impairment is reported in 56% of patients with unrepaired orofacial clefts involving isolated cleft palate, UCLP, BCLP and a lower prevalence in those with submucous clefts, of which 80% was in the age group 2 to 4 years of age which is also the crucial language-learning stage. (Zheng et al., 2009).

Speech defects are due to velopharyngeal insufficiency. This is when the soft palate is not able to make an adequate contact with the back of the pharynx to close off the nasal airway. It can also be secondary to poor hearing. Many individuals with treated

palatal cleft don't tend to have serious problems with communication. However about 25% fail to develop adequate speech. (Spriestersbach et al., 1973).

Dental anomalies in patients with orofacial clefts have been shown to occur more frequently than in healthy patients. These dental anomalies include abnormalities of tooth size and shape, enamel defects, delayed eruption, hypodontia and impactions.

- **Hypodontia** – 48% have missing lateral incisor on the cleft side compared to 6% on the non-cleft side in patients with UCLP (Tortora et al., 2008)
- **Supernumerary or supplemental lateral incisors (Tortora et al., 2008)**
- **Maxillary canine impaction** – patients with alveolar clefts have a twenty fold increase in risk (Russell and McLeod, 2008)
- **Impaction of first permanent molars**
- **Delayed dental development**
- **Hypoplastic teeth**
- **Microdontia**

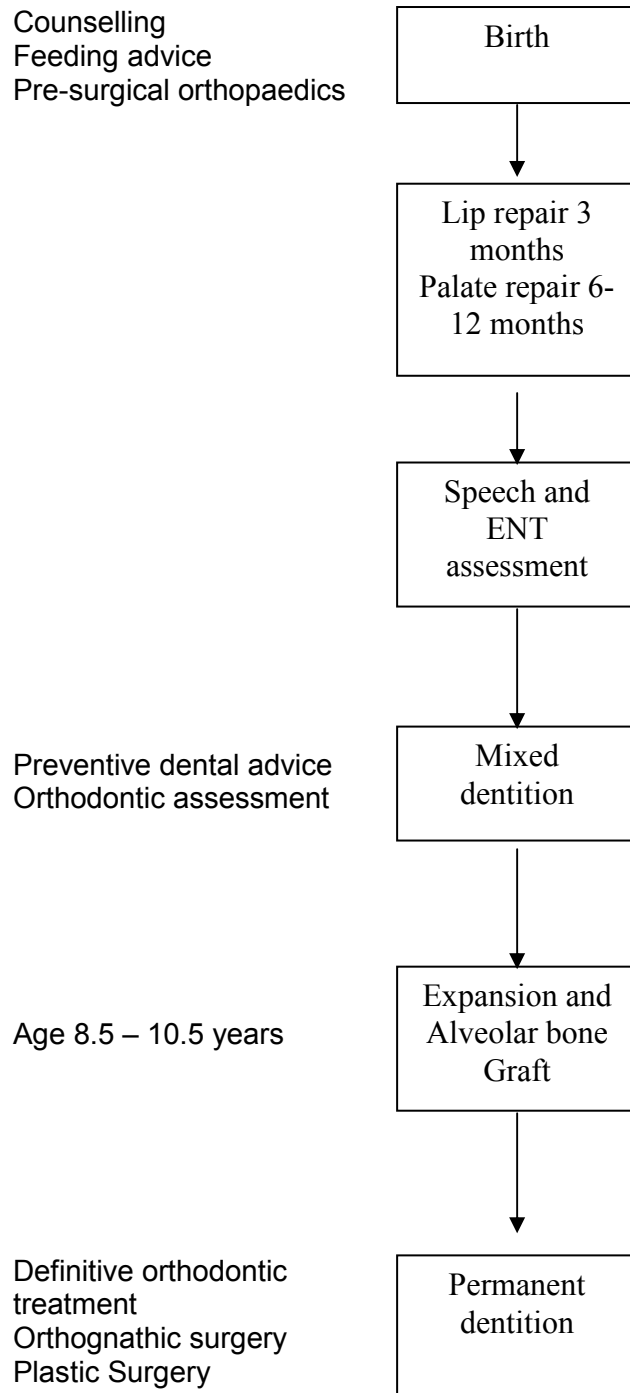
Deficient maxillary growth related to scarring of the maxilla following palate repair, this affects mid facial growth (Semb, 1991). This in turn may be reflected in the dental arch relationship (Mars et al., 1987) and therefore can manifest as anterior and/or posterior crossbite.

Psychology Literature suggests that although patients with cleft have normal intelligence and IQ, they are likely to have delayed development of expressive language. Research shows that parents quite often avoid speaking to them out of frustration of not being able to understand them. A positive attitude of family and friends appears to develop higher self-esteem than those with less family support (Cunningham, 1999).

It is generally agreed that patients with cleft lip only have fewer speech, language, hearing and dental problems than children with cleft lip and/or palate. Therefore this research will be based on patients with UCLP as it embraces all of these clinical problems. Although the clinical problems of patients with BCLP are more severe, it is also of a lower prevalence than UCLP and so more difficult to accrue an adequate sample size.

2.4 Current management of UCLP

Specialist treatment - from birth until 20 years of age



Birth

At the time of birth, the parents should receive counselling from a trained cleft health visitor. Contact details of a support group, cleft lip and palate association (CLAPA) is also a useful link. This is a voluntary group comprising parents of, and individuals with a cleft.

Special feeding bottles (e.g. Haberman feeder), which ejects fluid without the infant having to generate negative intra-oral pressure can be helpful if breast feeding is unsuccessful.

Presurgical orthopaedics with an upper removable/fixed appliance can be used to encourage lateral palatal shelf growth by stopping the tongue from sitting within the cleft site. The aim is to facilitate palate repair by approximating the cleft segments and aid feeding. This would be used until the time of palatal surgery. Recent evidence suggests that these devices offer no benefit to outcome either in terms of feeding (Prahl et al., 2005), nor does it reduce the incidence of maxillary collapse or facilitate initial contact between maxillary segments and so it has no positive benefit to surgical outcome. (Prahl et al., 2003).

Extra-oral lip strapping can be used in bilateral CLP to control growth of the premaxilla which facilitates lip closure, however it is not commonly used in UCLP.

Surgical Lip & Palate Repair

The surgical protocol for UCLP can vary between centres due to the lack of evidence to support any one protocol. Currently there is multi-centre randomised control trial (RCT) of cleft lip and palate surgery, evaluating the variation in surgical technique and staging. The aim is to assess whether there is a difference in outcome. It involves 10 centres and the study is being conducted as a parallel series of 3 RCTs of primary surgery for complete, unilateral cleft lip and palate (Semb et al., 1997)

Lip Repair

This is commonly performed at 3 months of age. It was suggested that lip repair in the neonatal period would result in a less of a scar. However, evidence shows that the results of neonatal repair are no better than later surgery and there is no psychological benefit to the parents (Slade et al., 1999).

Different surgical techniques are practised and all these techniques have a common aim which is to detach the muscles from its abnormal attachment and re-oppose them in their correct anatomical position. Dissection should be minimal to limit scarring that may hinder future maxillary growth.

In Third World countries, individuals manage to eat and speak just as well without repair of the lip, therefore it was concluded that lip repair is not necessary for function.

Palate repair

The aim is to separate the oral and nasal cavities with minimal effects upon normal growth and development. Palate repair is performed at 6-12 months of age using a single stage technique for combined hard and soft palate closure. There is evidence that good growth can be achieved by closure of soft and hard palate in a single stage at one year of age provided that the surgery is as atraumatic as possible (Zemann et al., 2007). This repair should not take place before 6 months as it puts individuals at risk of recurrent respiratory problems and severe restriction of maxillary growth.

Surgery to the palate has been linked to poor facial growth and so in some European centres this repair is delayed until 5 years of age or older in an effort to reduce the unwanted effects of surgery upon growth. However, late closure of the palate, does increase the risk of a poor speech outcome (Sell and Grunwell, 1990).

Hearing

Palatal surgery can improve Eustachian tube function. However, appropriate assessment and follow-up of hearing is necessary and other causes of deafness must be excluded. Follow-up every six months of all primary school aged children with cleft palate is recommended. If necessary, grommets are placed by ENT surgeons under local anaesthesia.

Speech Assessment

At 18 months each child will have had a routine assessment. This will be followed by further assessments using audio and video recordings where necessary at the following stages of life: 18months to 3 years, 3 years to 4.5 and then 5, 10, 15 and 20 years of age. The Speech and Language Therapist will contribute to diagnosis and treatment planning for children requiring further surgery for velopharyngeal insufficiency affecting speech outcome and symptomatic palatal fistulae.

Preventive Dental Advice / Paediatric Dentistry

As the deciduous teeth erupt, preventive dental advice (oral hygiene, dietary and the use of fluorides) is important to establish good oral health.

Paediatric dentistry and CLEFTSIS (see section 2.9) joined in 2002, where both look after children up till the age of 16. The aim is to maintain all primary and secondary teeth and to encourage dental compliance motivation for orthodontics.

Individuals are assessed at the following stages: 9-12 months, 2-5 years, age 7 for fissure sealants, 10 and 15.

The use of fluoride is practised according to the 'SIGN guidelines' (Evans, 2005) which recommend 1000ppm if < 5 years of age and 1500ppmF if > 5 years of age. Small smear if < 2 years of age and a pea sized amount if 2+ years.

Orthodontics

An average individual with UCLP will probably attend 60-70 orthodontic appointments up until the age of 17. The Eurocleft study (Shaw et al., 1992b) shows that increasing appointments with orthodontists does not conform to better outcome. Clinical Standards Advisory Group (CSAG) report also recommends that all orthodontic treatment should be carried out by the orthodontic cleft specialist (Sandy et al., 1998) Treatment should be limited to discrete episodes such as:-

- Presurgical orthopaedics (from birth)
- Preparation for alveolar bone grafting (7+ years)
- Align the maxillary dentition (usually using fixed appliances) if the appearance causes the child distress or the irregular teeth are traumatising soft tissues (6+ years)
- Definitive alignment of the maxillary and mandibular teeth using fixed appliances (12+ years)
- Decompensation and alignment for orthognathic surgery using fixed appliances and detailing the occlusion post-operatively (late teens/adult)
- Fabrication of obturators, palatal lift appliances and electropalatography (EPG) appliances.

Preparation for alveolar bone grafting (ABG)

This is usually undertaken between 8.5 and 10.5 years when the root of the maxillary canine is half to two-thirds formed. The decision to treat early at 8.5 years would be to accommodate an unerupted upper lateral and the later date of 10.5 years would be to accommodate an unerupted upper canine.

The purpose of alveolar bone grafting is to provide alveolar bone, usually taken from the iliac crest, for canine eruption and into which teeth may be moved or implants placed. It also supports the alar base of the nose and closes residual palatal fistulae. Before the alveolar bone graft, supernumerary, deciduous and any other teeth with poor long term prognosis should be removed.

Expansion, with a quad/tri-helix, is often necessary before bone grafting to expand the collapsed cleft segment and improve access to the site for the surgeon.

Concurrently, fixed appliances can be used to align anterior teeth and correct anterior crossbite. Due to the limited teeth available a simple fixed appliance on central incisors and first molars can be utilised. It is important to correct an anterior crossbite at 9 years of age as it is more difficult later as a result of differential mandibular growth. Care should be taken not to move teeth towards the cleft as the lack of bone may cause root exposure. Often the aim is to accept the mesial-distal tip and rotations present in the upper incisors. Therefore, when placing the brackets it is wise to accept the inclination of these teeth rather than try to upright them. After the bone graft the brackets can be replaced and the roots moved into the correct position. After expansion, a transpalatal arch with palatal arms, is fitted which, retains the expansion and allows surgical access.

After surgery the patient is reviewed three months later and a radiograph is taken to assess the graft. The bone graft needs to be loaded with orthodontic treatment or natural eruption of teeth, otherwise unloaded bone will resorb which hinders implant placement in the future.

In summary the orthodontist needs to expand, retain, align the upper centrals with cautious movement of the roots and remove supernumeraries prior to alveolar bone grafting.

Definitive alignment of the maxillary and mandibular teeth using fixed appliances.

Where there is no appreciable skeletal deformity, definitive alignment of the maxillary and mandibular teeth is carried out using fixed appliances when the permanent dentition is established.

Orthognathic surgery

Where there is an appreciable skeletal discrepancy, which may require orthognathic surgery, it is wise to assess the patient on a joint clinic prior to planning any definitive alignment. Orthodontic decompensation can take up to 2-3 years of treatment. Upper arch fixed appliance treatment may be undertaken just to align the upper teeth prior to orthodontic decompensation, but only if the patient wishes as it is easy to lose patient compliance at an early stage.

Maxillary hypoplasia is a common developmental problem in cleft lip and palate deformities and normally results from a combination of congenital reduction in midfacial growth and the effects of surgical scar from cleft palate repair. Surgery with resulting scars seems to be recognised as the main reason for poor growth of the maxillary region (Mars and Houston, 1990). The treatment aim would be to advance and down graft the maxilla for optimal dentofacial aesthetics and function. Patients with cleft need greater advancement than patients without cleft, however the palatal scar resists the ability of the maxilla to be advanced by a large amount and it contributes to the high risk of relapse compared to non-cleft class III patients. Furthermore, during maxillary advancement by a Le Fort 1 osteotomy, the soft palate is pulled forward, increasing the anteroposterior distance from the posterior pharyngeal wall. This change may affect the velopharyngeal closure function and hence affect speech. (Cheung and Chua, 2006). Therefore it is important that velopharyngeal function is assessed prior to surgery.

Consequently patients are advised that surgery may be required in both jaws. The use a mandibular setback osteotomy to compensate for the extent of required

maxillary advancement. Surgery should be deferred until the growth rate has slowed to adult levels, which is approximately age 15 for females and 17 for males.

Distraction osteogenesis It is hypothesized that distraction osteogenesis may help reduce the adverse effects associated with Le Fort 1 osteotomy when applied to correct maxillary hypoplasia. The surrounding soft tissues may have a better chance to adapt to the skeletal changes through slow movement than with the sudden changes produced by conventional Le Fort 1 osteotomy. A recent meta-analysis of cleft maxillary osteotomy and distraction osteogenesis, infers that there is no conclusive data on any differences in surgical relapse, velopharyngeal function and speech between the two techniques and that both can deliver a marked improvement in facial aesthetics (Cheung and Chua, 2006). Most surgeons would decide to distract when the maxilla needs at least 7mm of protraction. This helps to reduce relapse as it produces skeletal change by slow movement that produces gradual stretching of the soft tissues. After placement of the distractor, there is a latent period of 3-4 days after which it is activated 1mm per day for 3 months. The decision to distract instead of Le Fort 1 osteotomy is made when the maxilla needs at least 7mm protraction.

Hypodontia – replacement of missing teeth with implants can also be undertaken at the end of growth assuming that there is adequate bone volume at the site of implant placement.

Secondary/Plastic Surgery

This is voluntary to improve nasal aesthetics, lip revision and close residual palatal fistulae. Velopharyngeal insufficiency can be corrected following maxillary

advancement and rhinoplasty should be carried out after orthognathic surgery as movement of the underlying bone will affect the contour of the nose.

Psychology

The CSAG recommendations (see section 6.2) suggest that every cleft centre should have an appropriately trained psychologist as a core member of each cleft team. Parents need reassurance and support at birth as their emotions are compared to those of a grieving process. Psychological support may be required during primary and secondary schooling when children start to notice that they are different and may be teased (Shaw, 1980) found that dental features were the fourth most common cause of teasing, harassment and a target for nicknames. During adolescence these patients will also be undergoing orthodontic treatment which will not aid their self-esteem and research shows that as adults they are more prone to depression, anxiety and suicide (Christensen et al., 2004). Therefore, psychological care from birth to adulthood is necessary for these patients.

Traditionally, World Health Organisation (WHO) have classified disease according to morbidity. However recently, they have introduced another classification according to function as well. The consequences of UCLP are related to morbidity and function. However, the incidence of a functional disability would be greater than morbidity. This is reflected by the number of individuals with UCLP in the developing world who have either adapted or accepted their functional disability; due to the lack of resources and /or personal financial constraints (www.who.int/classifications/en).

2.5 Standards of care

To date, there is not an agreed gold standard for surgical technique nor timing for repair of UCLP. Randomised trials would be the best research method for assessing the effectiveness of cleft care, thus a means for improving standards of care.

There have been two small (20 patients in each arm) randomised trials of primary surgery conducted in 1960s, which investigated two different surgical techniques and their effect on facial growth. Much of the literature consists of retrospective studies or is at the level of case series and case reports (Shaw et al., 2001). The relative infrequent occurrence of UCLP and so insufficient numbers for random allocation is difficult to achieve from a single centre.

In the 1980's and 1990's intercentre comparison studies were conducted in the UK and Europe. It revealed that the difference between the best and the worst centres differed by a factor of 8 in terms of their need for maxillary osteotomy (Shaw et al., 2001). Furthermore, care in the worst units was much more complex and expensive.

2.5.1 A Six Centre European Comparison Study : Eurocleft Study.(Shaw et al., 1992a, Molsted et al., 1992, Mars et al., 1992, Asher-McDade et al., 1992, Shaw et al., 1992b)

This was a prospective, multicentre clinical audit of treatment outcome for complete unilateral cleft lip and palate between six European cleft centres. A cohort of subjects aged between 8-10 years were followed through to age 17. The dental arch relationship, craniofacial form and nasolabial appearance was assessed from study models, cephalometric radiographs and photographs, respectively. These records were collected at different time points; age 8-10, 12 and 17. Improved midfacial development was found to be more severe at some centres compared to others. Few if any centres conformed to the same approach in surgical technique and timing. The centres with the best surgical outcomes were associated with standardised, centralised and high volume operators, where at least 30 patients were treated per

clinician. Neither of these centres employed the use of presurgical orthopaedics as it was found to be of no value to the dental arch relationship and craniofacial morphology. The centres with poor surgical outcomes were associated with multiple operators, low caseloads and inconsistent protocols. It found that the two UK centres that participated were the weakest on almost every aspect of care.

2.5.2 Eurocleft project 1996-2000 (Shaw et al., 2001) & (Shaw et al., 2000)

Differences in outcome and treatment protocols revealed by the European intercentre comparison study, instigated a stimulus for improvement in services across Europe. Subsequently, a project was funded by the European commission with the aim of raising standards across Europe and to establish a network to promote research and quality improvement programmes. It involved 201 centres within 30 European countries. Each centre was invited to register and was asked for details of professional team, service organisation and clinical protocols. It found that 194 different protocols were practised within 201 centres.

As a result of the findings from the intercentre comparison of six centres, it was proposed that each centre adopts centralised care where each clinician has an annual caseload of 40 patients. This would also generate an appropriate caseload for audit and research purposes. In order to participate in comparative studies the European consensus proposed a common protocol on record collection across Europe. A minimum of four study models should be taken as follows

- (a) primary surgery
- (b) at 5/6 years
- (c) at 10 years and
- (d) at 18+ years of age.

The protocol suggests that all study models are prepared to a standard manner, in order to allow a fair comparison. The impressions must include all the teeth, the palate and the buccal sulcus.

The precise preparation of standardised study models is illustrated in Chapter 4: Method and Materials (see Figure 10)

A collaboration of 201 centres, all conforming to a standardised care of treatment enables a multicentre randomised trial to take effect. This has instigated a prospective trial (Semb et al., 1997). This is a randomised clinical trial of complete unilateral cleft lip and palate surgery and involves 8 Nordic and 2 UK teams. This study is evaluating the difference in outcome by varying surgical technique and staging, where the usual local method is being compared with a common protocol (lip & soft palate closure at 3-4 months, hard palate closure at 12 months). This investigation is designed to compare the relative merits of four surgical methods. The study is ongoing and the sample size has been calculated such that 280 infants need to be recruited into the study over a three-year period.

2.5.3 Clinical Standards Advisory Group (CSAG, 1998)

The findings from the six centre European study was a concern in the UK, which led to the establishment of CSAG national investigation into cleft care in the UK. This study reported upon clinical outcomes in a total of 457 subjects, 5- and 12- year - old children affected by non-syndromic unilateral cleft lip and palate.

The study showed that there were 57 active cleft teams in UK with 75 surgeons undertaking primary repairs. Three of these surgeons had less than 5 repairs per annum. One hundred and five orthodontists were involved in the care of these children. Seventy per cent claimed to have had active databases but more than half were unable to provide evidence that the information from databases was accurate. Clinical information (photos, radiographs, models) was collected for clinical

management purposes only. Surgical procedures were not accurately documented to allow inferences to be drawn about the relationship between interventions and outcomes.

The dental arch relations were measured with the GOSLON and 5- year-old indices, with 37-39% of both age groups were either 'poor' or 'very poor'. Seventy per cent of the 12-year-old patients had a Class III skeletal relation and 42% of bone grafts were seriously deficient or had failed. Basically, it concluded that neonatal surgical outcome for patients with UCLP in UK were poor. Furthermore, the training of appointed consultant orthodontists involved in cleft care was questionable. As result of these findings, CSAG made a number of recommendations:

- cleft care should be centralised, with expertise and resources concentrated in 8 to 15 national centres.
- a common nationwide database should be established for all patients with cleft
- minimum record collection at specified time points, which will further facilitate multicentre audit.
- training for specialist cleft clinicians should only be provided in cleft centres where high-volume and high-quality clinical experience was available.
- an annual caseload of 40 patients for each cleft clinician.
- purchasers and commissioners were asked to purchase cleft care only from centres of excellence

2.5.4 The importance of measuring surgical outcome

A lot of funding and planning has been invested into changing this service to a centralised service, where the main focus is for a highly trained multidisciplinary team to treat a high volume of patients to a standardised protocol, where meticulous documentation is exercised. The protocol also recommends long term follow up and audit in order to enter into comparisons with other centres.

Therefore the need for an accurate measuring tool to measure standards of care is ever more pertinent in light of recent changes which, is focused on improved care. It is also believed that early predictors of treatment outcome in cleft care are timely and relevant because they provide a means to reduce the length of research studies without increasing the sample size (Roberts et al., 1991).

Patients with UCLP begin treatment at just a few months of age, which involves surgery. The result of this surgery determines their future facial and dental appearance. There is strong evidence to suggest that impaired growth of the mid-face is related to the effects of primary surgery in infancy (Mars and Houston, 1990). Recent studies indicate that poorly performed primary surgery is likely to compromise facial growth, dental development (Mars et al., 1992) and speech (Wyatt et al., 1996). Surgical improvement is a necessity in the field of evidence-based medicine where patient care is based on national and international standards.

Standards of treatment should be assessed regularly in order to sustain a high level of care, especially as successful early treatment is essential to later stages of facial and dental development (Shaw et al., 2001).

2.6 Indicators of & methods for assessing surgical outcome

2.6.1 Dental arch relationship

Compromised facial and maxillary growth is a common finding in many patients with repaired complete clefts of the lip and palate. It is generally agreed that mid-facial growth is an indicator of surgical outcome, since it appears that primary surgery used for the correction of the lip and palate defect has some effect on the maxilla (Semb, 1991). However, some claim that maxillary hypoplasia is an intrinsic primary defect, while others believe it to be secondary to surgical repair (Bishara, 1973). A study by (Mars et al., 1990) supports the view that there is potential for normal facial growth in a patient with UCLP and that palatal closure is likely to cause maxillary hypoplasia that may be severe. This was shown by comparing 3 separate subgroups of Sri Lankan subjects against a control group of healthy, noncleft Sri Lankan subjects, all over 13 years of age in each group. The three separate subgroups were: (a) those who had totally unrepaired cleft lip and palate, (b) those who received lip repair in infancy but not palatal repair, and (c) those who had lip and palate repair in infancy. The three subgroups accounted for 60 male subjects and the control group had 23 male subjects. The results showed that subjects who had no surgery had a potential for normal growth. Subjects who had lip repair in early infancy showed relatively normal maxillary growth, but maxillary hypoplasia was common when the palate had also been repaired early.

Therefore the outcome of primary surgery in UCLP is a crucial factor in the subsequent development of the maxilla and facial appearance.

The problems of growth of the dentofacial complex are generally reflected in the antero-posterior dental relationships (Noverraz et al., 1993). The AP dental relationship is considered to be of the most clinical importance, whereas vertical and transverse relationships are less critical and primarily help in ranking borderline

cases (Chan et al., 2003). The antero-posterior dental relationship is reflected in the dental arch relationship.

Many studies have used the dental arch relationship to evaluate the common outcome of cleft palate surgery. Poor surgical outcome tends to result in constriction of the maxilla and therefore success or failure can be related to the dental arch relationships and the frequency with which crossbites occur ((Mars et al., 1987); ((Mars and Houston, 1990).

It is suggested that the measurement of overjet, while being quick and simple compared to the GOSLON Yardstick, could rank the casts comparably, and would give results that allowed parametric statistical analysis. It was therefore suggested that the measurement of overjet be considered as an alternative indicator of outcome in the management of UCLP (Morris et al., 1994). However, the use of overjet over other indices, such as the modified Huddart Bodenham Index, which is also quick and simple to use and lends itself to parametric statistical analysis, has not been validated. Therefore until further evidence, crossbites should be used to assess dental arch relationship.

2.6.2 Indices

'In the orthodontic context, an *index* is used to describe a rating or categorising system that assigns a numeric score or alphanumeric label to a person's occlusion.'
(Shaw et al., 1995)

Indices have been developed to measure treatment effectiveness relating to different aspects of anatomic form and function in the parts affected by the clefting process.

Ideal measure of outcome should be easy to learn, quick to apply and have good reliability and validity.

Reliability – is the ability to reproduce the original ratings or scores when the subject is re-examined by the same or different examiners.

Validity – is the extent to which an instrument measures what it purports to measure.
(Shaw et al., 1995)

World Health Organisation (WHO) recommendations for an ideal index

According to WHO, the general requirements of an index are as follows (World Health Organisation 1977)

- 1) It should be reliable
- 2) It should be valid
- 3) It should be acceptable by the profession
- 4) It should require minimal judgement
- 5) It should lend itself to statistical analysis
- 6) It should be administratively simple

The Eurocleft and CSAG studies have demonstrated that it is possible to detect differences in outcome using indices to assess dental arch relationships.

2.6.3 Antero-Posterior Skeletal Bases

The evaluation of the antero-posterior relationship of skeletal bases using hard and soft tissue outline of these structures can be considered a useful outcome of cleft lip and palate. The only means of doing this would be radiographs such as a lateral cephalogram which provides both hard and soft tissues. Profile photographs can also be used to evaluate soft tissues.

2.6.4 Lateral Cephalometry

A study by (Mackay et al., 1994), shows that evaluating the antero-posterior skeletal bases in 5-year-old subjects with UCLP from a lateral cephalogram proves to have

many errors, thus rendering this method unjustifiable as a basis for measuring the outcome of surgery. These errors include taking standardised lateral cephalograms, processing and tracing errors.

The study also concluded that the accuracy of measuring the antero-posterior position of the maxilla for a 5-year-old child on a cephalogram is questionable. This is because it was difficult to identify anterior nasal spine, subspinale and pterygomaxillare on the lateral cephalometry. The anterior surface of the maxilla was frequently found to be convex in contour and in over half of the radiographs, the unerupted maxillary central incisor appeared rotated, thus confusing the anatomy of this area even further.

2.6.5 Soft tissue profile

Examination of soft tissue profile of individuals with orofacial clefts may indicate the extent of maxillary hypoplasia, thus the impact of treatment on anterior-posterior growth.

2.6.6 Photographs

This method was used in the Eurocleft study, where photographs were scored by a panel after the photographs had been masked, to eliminate the confounding effect of other features of facial appearance which have been shown to affect judgement (Shaw, 1981). The study found that acceptable levels of reliability and reproducibility could be obtained with this methodology, although reliability was poor compared with scores for dental arch relationships (Asher-McDade et al., 1992).

Another study also assessed the reliability of this method and found that panel assessments had poor reliability, and were liable to errors because of their subjective nature. The scoring system comprised of a five-point scale 1) excellent 2)

good 3) satisfactory 4) poor 5) very poor, similar to the Eurocleft scoring system. Both studies found that extreme ranges of scores were rarely applied and that the bulk of scores was clustered around the midpoint scale. The study concludes that the reliability of this system is acceptable and that it needs to be a more objective assessment of the profile (Bearn et al., 2002)

Also, examination of soft tissue profile provides only limited understanding of the relationships of structures in the midline and cannot quantify aberrant anatomical position or asymmetry (Molsted et al., 1992). Three-dimensional (3D) nature of the cleft malformation is a prerequisite for quantifying the magnitude of the anomaly and measuring change following surgical repair.

2.6.7 Stereophotogrammetry

Soft tissue features of patients with UCLP are significantly different to patients without cleft (Hood et al., 2004). Therefore soft tissue features can be a source of measure for surgical outcome. This was evaluated using stereophotogrammetry, which is an extension of photogrammetry.

This produces a 3D depth map of the face, which is created from images captured by a triangular camera arrangement. The system digitises the facial surface and provides a high resolution, full colour 3D virtual face. This is a permanent record of the subject and can be interactively viewed and manipulated on a computer screen. A study by (Hajeer et al., 2004) showed good accuracy of digitising landmarks and used 20 landmarks in their study that had been shown to be highly reproducible with an error of less than 0.5mm.

Stereophotogrammetry is a user-friendly system for quantitative evaluation of soft tissue features which contribute to facial asymmetry in patients with cleft lip and /or palate. Up until now, collecting 3D information in infants has not been possible because imaging techniques have required sedation, ionizing radiation or demanded

a level of co-operation that young infants and children are not capable of achieving. It is a non-invasive system which captures facial morphology in 50 milliseconds (Hood et al., 2004). This makes the system highly compatible to use for 3-month-old infants, enabling measurements to be taken as early as 3 months providing surgeons a source of immediate feedback, following surgical correction to assess outcome and subsequent re-evaluation of surgical rationale. Therefore early objective evaluation will help develop even better surgical management practices to ensure excellent results in the long-term.

The use of stereophotogrammetry also helps with surgical planning as it provides data of the non-cleft side, so that during the operation, these measurements can be used as a reference. A paucity for population norms for face shape in children in the UK, makes it difficult to compare cleft data with normative data.

Stereophotogrammetry has been used to quantify the relative contributions of the nose and lips to overall facial asymmetry in 3-year-old, 9-year-old and 1-year-old infants, post surgery. The study found that lip symmetry in UCLP infants appears to improve after surgery, proving that this method has the ability to detect findings very early in an infant's life.

Therefore, soft tissue features are an another important source of surgical outcome, other than dental arch relationship. Both are evaluated using different methods, however, compared to study models, it is an expensive tool and not as simple to operate.

2.6.8 Cephalometric radiography

Soft tissue contour is visible on the radiographs and so it would be possible to obtain a more reliable picture of facial form using a soft tissue analysis. However, it would prove difficult to standardise cephalometric radiographs because of the difference in

x-ray magnification between centres. Furthermore, cephalometric radiography for growth studies, particularly in young patients, is questionable.

2.6.9 Palatal Height

It was suggested that palatal height and anterior maxillary arch depth could be considered as another possible indicator of surgical outcome. However, its relevance to any functional consequence is uncertain (Kharbanda et al., 2002)

2.7 The choice of study models for outcome assessment

Study models provide an inexpensive, easy and minimally invasive 3-dimensional record of the dental arch relationship. They are the most universal method of recording surgical outcome and therefore the most readily available of all records. Furthermore, standardised study models provide a base for universal comparison. As all the study models in the CLEFTSIS database are trimmed to standards stipulated by Eurocleft, a like for like comparison can be made between all the dental arches.

There are a couple of studies which support the use of two-dimensional digital photographs of study models, (Nollet et al., 2004) & (Ali et al., 2006). Both were carried out on 9-year old and 5-year old study models using the 'GOSLON Yardstick' and the 'modified Huddart Bodenham' indices, respectively. They both concluded that the use of two-dimensional photographs is a consistent, reproducible method for rating dental arch relationships of patients with UCLP.

However, it was decided not to utilise this method as it has not been validated for the 'EUROCRAN Yardstick' and so it would not be deemed a fair comparison.

Another option would have been to extract outcome measures from three-dimensional virtual study models as it enables the observer to rotate the study models in all directions on the screen. This would enable a more accurate assessment compared to two-dimensional photographs as any small discrepancies could be missed with the latter, giving a false outcome measure. However, this method could not be utilised as none of the indices have been validated for three-dimensional study models. A study to evaluate this method is in progress at another dental teaching centre in Scotland (Personal communication)

Therefore, the evidence thus far suggests that it is possible to detect differences in surgical outcome as early as 5yrs of age, by assessing the dental arch relationships on study models, using indices.

2.8 Different types of indices

A number of methods, which measure maxillary arch constriction have emerged. These include: GOSLON (Mars et al., 1987); the 5-year-old index (Atack et al., 1997a); the modified Huddart Bodenham (Mossey et al., 2003) (Gray and Mossey, 2005); the EUROCRAN Yardstick (Oskouei, 2007). Each will be discussed below, together with their advantages and disadvantages.

2.8.1 GOSLON

Great Ormond Street, London and Oslo (Goslon) index, introduced by (Mars et al., 1987) is a standardised ranking system specifically developed for categorising dental arch relationships in children with UCLP. Unlike other systems, the GOSLON yardstick is treatment related (e.g. anterior crossbite with retroclination of the incisors can be corrected more easily than anterior crossbite with normal incisor inclination) and is therefore more informative than a crossbite score alone. Although this scoring system is based on the potential for orthodontic correction, this is influenced not only by surgical outcome but also by inherited skeletal pattern.

The system was developed for categorising dental arch relationships in 10-year-old children with UCLP examined in the late mixed or early permanent dentition (Mars et al., 1987).

It categorises malocclusions in patients with UCLP according to anteroposterior (AP) arch, vertical labial segment and transverse relationships.

Scoring system

A score of 1 means a favourable AP relationship for orthodontic correction and a score of 5 means a very poor AP relationship with osteotomy necessary for correction. A score of 3 usually means an anterior end-to end situation.

The 5 categories are:-

Group 1 (excellent), at a very favourable end of the scale, represents the most advantageous skeletal form with a positive overjet and overbite. Patients in this category typically exhibit an Angle Class II div I malocclusion, which in the case of operated cleft lip & palate patients, represents a rare and beneficial relationship.

Group 2 (good) is also a favourable relationship that can be corrected with straightforward orthodontic treatment.

Group 3 (fair) presents an edge-to edge dental relationship that will require more complex orthodontic therapy to correct the Class III malocclusion and other possible arch deformities, but a good result can still be anticipated.

Group 4 (poor) include relationships at the limits of orthodontic treatment that because of unfavourable facial growth, may necessitate an orthognathic procedure.

Group 5 (v poor) represents a significant skeletal Class III relationship with obligatory surgical correction.

The categorisation is based on the collective views of a panel of experienced orthodontists and so it is based on the consensus opinion that reflects the features of good or poor occlusion for repaired UCLP.

Advantages

The major advantage of the GOSLON index is that it considers clinically important variables in all 3 planes of space and permits the ranking of models in order of difficulty to achieve a favourable outcome.

It has been shown to have good inter and intra examiner reliability (Mars M, 1985).

Disadvantages

10 years of age

There has been a move towards earlier bone grafting to allow eruption of permanent teeth (2's & 3's) so its unlikely that future 10 year old samples will all have subjects similarly matched as far as treatment received is concerned. They may have had orthodontic treatment as well and so these models of 10 year old patients will not allow measurements to reflect the outcome based on surgical intervention alone.

It was therefore suggested that the GOSLON yardstick could be applied to 5 year old study models but a study by Attack et al (1997b) showed that this was not applicable. The study assessed the reproducibility of the GOSLON yardstick on 5 year old study models and found that the agreement was at worst, moderate and at best, very good (Altman, 1991).

However another study by Noverraz et al 1993 concluded that the GOSLON yardstick is a useful assessment of dental arch relationships at all stages of dental development i.e. deciduous dentition, early mixed dentition, late mixed dentition and permanent dentition (age range 4-17) thus concluding that the GOSLON yardstick was suitable for longitudinal research. Interestingly, the study quotes that 'GOSLON scores for all stages of dental development was adequate. This was surprising as the kappa scores were on average 0.8 which equates to very good agreement and hence one would have expected the study to deduce that the 'GOSLON yardstick' is an accurate tool for the assessment of dental development from age 4-17. The study did highlight that the early mixed dentition (mean age 8 years) was the most difficult period to categorise with this index. Therefore both studies arrive at different conclusions.

Categorical

The GOSLON yardstick is essentially a subjective, ordered, categorical classification, which is likely to be less powerful than an objective continuous numerical

measurement scale. Furthermore, a continuous scale measurement more readily satisfies the assumptions of parametric statistical analysis.

Calibration

The GOSLON yardstick requires the judges to be trained in the use of this index and recalibration is necessary to ensure consistency. 10 reference models have to be used for comparison during the assessment and scoring of study models of patients with UCLP. All of which adds to the complexity of the exercise.

Not versatile

The GOSLON yardstick can only be used to score UCLP and no other cleft types.

Validity

The validity of the GOSLON yardstick has not been investigated and it is recognised that this is difficult since it requires a cohort of adults with unilateral cleft lip and palate treated with only primary surgery.

Comparing the adult and 10-year-old's dentition to see if GOSLON rating at 10 years of age gives a valid prediction in the adult dentition would enable a valid assessment. However, this 10 year old sample which has been treated with only primary surgery very rarely exists as most patients at this stage have undergone orthodontics, restorative procedures and usually alveolar bone grafting (Shaw and Semb, 1990) all of which can mask the effects of primary surgery.

A study by (Noverraz et al., 1993), suggested that the power of the GOSLON Yardstick is in its face validity, which is said to apply when the relevance of a measurement appears obvious to the investigator.

Therefore it is assumed that true validation of the index is not possible and it relies on face validity.

2.8.2 5-year-old index

This index assesses study models of 5 year olds and was developed to overcome the shortcomings of the GOSLON yardstick (Atack et al., 1997a). It allows surgeons to assess their outcome more accurately so that they can make appropriate changes to their clinical skills.

This index is also categorical and has 5 categories:-

Group 1 (excellent)	positive overjet with average inclined or retroclined incisors. No crossbites Good maxillary arch shape and palatal vault anatomy
Group2 (good)	positive overjet with average inclined or proclined incisors Unilateral crossbite/crossbite tendency +/- open bite tendency around cleft site
Group 3 (fair)	edge-to-edge bite with average inclined or proclined incisors;or reverse overjet with retroclined incisors. Unilateral crossbite +/- open bite tendency around cleft site
Group 4 (poor)	reverse overjet with average inclined or proclined incisors Unilateral crossbite,+/- bilateral crossbite tendency +/- open bite tendency around cleft site
Group 5 (very poor)	reverse overjet with proclined incisors Bilateral crossbite Poor maxillary arch form and palatal vault anatomy

This index has been shown to have excellent and good intra- & inter-examiner reliability, respectively.

Advantages

The 5-year Index is therefore a more reliable tool in measuring 5-year-old models than the GOSLON index (Atack et al., 1997b).

Disadvantages

True validation of this index is not possible and it relies on face validity (Atack et al., 1997b).

Just like the GOSLON index, the 5 year old is also ordinal, not versatile and the examiners need to be calibrated, therefore making it just as complex to use for scoring surgical outcome.

2.8.3 Huddart Bodenham (1972)

The modified Huddart Bodenham Index was derived from the original Huddart Bodenham Index, in order to make it more versatile.

The original Huddart Bodenham could only be applied to the following:-

- deciduous teeth
- patients below age 6
- UCLP
- It has 5 categories for scoring incisors and 3 categories for scoring canines and molars.

In contrast, the modified Huddart Bodenham, can be applied to:-

- both deciduous and permanent teeth
- patients at any age above 3
- any cleft type
- Again it has 5 categories for scoring incisors, canines and molars, making it much easier to use.

The original Huddart Bodenham index was devised following the evaluation of 2 other categorical indices, devised by (Pruzansky and Aduss, 1967) and (Matthews et al., 1970), both of which measure the presence and degree of crossbite both anteriorly and posteriorly. The study (Huddart and Bodenham, 1972) concluded that both these categorical indices were not reliable in the hands of different observers, because categories encompass sharp delineation and sharp delineation does not extend to occlusion. Therefore, a great deal of subjective judgement needed to be employed, rendering the indices unreliable in different hands as it would be very difficult to establish a common assessment criteria. The paper also stated that by employing two different indices, which were so dissimilar, made effective comparison of results between centres very difficult. Therefore the authors attempted to devise an index which was numerical, give more detailed information and lend itself to statistical analysis.

Pruzansky and Aduss, 1967 divided the occlusion into six categories:

- 1) no crossbite**
- 2) canine crossbite only**
- 3) buccal crossbite only**
- 4) anterior and buccal crossbite**
- 5) anterior and canine crossbite**
- 6) incisor crossbite only**

In contrast, Matthews et al, (1970) divided the occlusion into:

Class A – where all the segments of the maxilla are in normal occlusion with the mandible.

Class B (1) – the tooth bordering the cleft on the lesser segment is in lingual occlusion.

Class B (2) – normal occlusion of the greater segment but lingual occlusion of the lesser segment

Class B (3) – the maxillary arch is perfect but is too small.

Class C – an overall Class III occlusion of all segments of the maxilla, and, in addition, collapse of some part of the small maxillary arch.

2.8.4 The Modified Huddart Bodenham Index

This scoring system overcomes all of the above disadvantages. (Mossey et al., 2003) and (Gray and Mossey, 2005) compared the reliability and reproducibility of this index with the GOSLON and 5- year old indices. It showed the modified Huddart Bodenham Index to be more reliable, objective, sensitive and simple to use. Furthermore it is more versatile because it is applicable to any age from 3 years upwards and any cleft type.

The scoring system uses the frequency and severity of crossbites where each maxillary tooth is scored according to its relationship with the corresponding tooth in the mandible. Individual scores are added together to give a total score. The more negative the score the more the arch constriction.

This system is described as 'modified' because it was developed from the original 'Huddart Bodenham' index, developed in 1972 (Huddart and Bodenham, 1972).

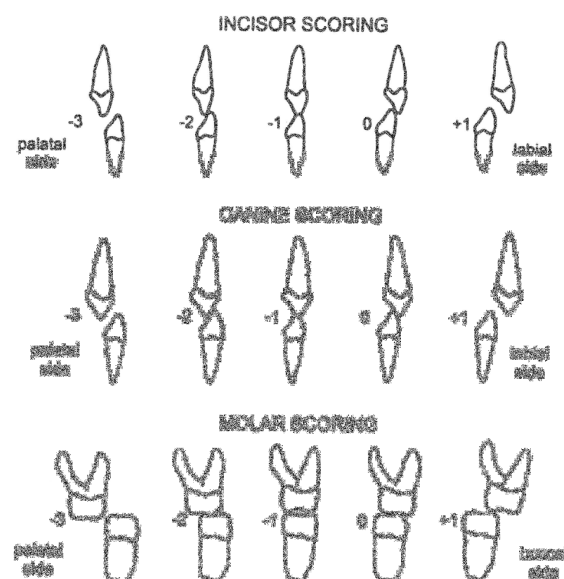


Figure 9: Pictorial chart for modified Huddart Bodenham, adapted from (Heidbuchel and Kuijpers-Jagtman, 1997)

Advantages

Reliability & Objectivity

The study by Gray and Mossey (2005), found the modified Huddart Bodenham Index to have more intra- and inter-rater reliability than the GOSLON & 5-year indices. This study also showed that the modified Huddart Bodenham Index is an objective scoring system as the scorer does not require clinical judgement or experience. Mars et al (1987) stated that in some cases precise allocation to a GOSLON category may be ambiguous, which in turn will adversely affect intra- and inter-rater reliability. Therefore the more objective the index the more accurately it can reflect the extent of interarch discrepancy.

Simplicity

No calibration course nor reference models are required, which eliminates the need to train examiners and improves intercentre collaborative studies. Scores can be calculated very quickly and consistently.

Versatility

The modified Huddart Bodenham Index can be used for any age and any cleft type (Tothill and Mossey, 2007). In contrast, GOSLON and 5 –year-old Index are restricted to 10 and 5 year old age groups respectively and can only be applied to UCLP.

Sensitivity

The modified Huddart Bodenham Index is a continuous rather than a categorical scale measurement and so provides a greater degree of sensitivity and also satisfies the assumptions of parametric statistical analysis. The more sensitive the index, the greater its ability, to detect an interarch discrepancy. Therefore the modified Huddart Bodenham Index has the ability to differentiate the severity within the categories that

would be identified by the GOSLON or 5-year-old indices. This has been demonstrated by the following studies by Gray and Mossey (2005) and Weir (2009).

Disadvantages

Validity

The modified Huddart Bodenham scoring system has been validated on study models by Mossey et al (2003) and Gray and Mossey (2005). Validation was carried out against 10 GOSLON and 10 5-year-old index reference study models. These reference models represented the 5 categories in their respective scoring systems and therefore these models were an ideal source to validate the modified Huddart Bodenham Index. This was a pilot study which showed that there was high correlation of the modified Huddart Bodenham Index with the GOSLON and 5-yearold index, which showed that the modified Huddart Bodenham Index measures what it is meant to measure and hence is valid. The validity of the modified Huddart Bodenham Index was further evident in a larger study by Gray and Mossey (2005) which involved a larger sample size of 100 patients. This same study showed that the modified Huddart Bodenham Index to be reliable and objective.

2.8.5 EUROCRAN Yardstick (Oskouei, 2007)

This index was developed by the participants in the EUROCRAN project (2000 - 2004). This project was a continuation of the EUROCLEFT project with the aim to improve research capabilities.

This index was developed after findings from assessing a mix of 118 cases from different European centres. These cases were scored using the GOSLON Yardstick and 5-Year-Old indices. The scores showed that only one of the cases was graded as 5 and two cases as 1 by all the examiners involved in the study. Therefore due to a lack of use of the extremes of scales from 1 to 5, it was decided to reduce the grades option from 1 to 4 in antero-posterior, vertical and transverse dimensions,

instead of the 5-grading scale. In addition, a 3-grade scale is allocated for rating the palatal form.

Therefore the EUROCRAN Yardstick is a modification of the GOSLON Yardstick and 5-Year-Old indices and it is also designed to evaluate surgical outcomes in patients with UCL/P. It is applied to study models and the major components are the dental arch relationship in the antero-posterior, vertical dimensions and the palatal form.

The EUROCRAN Yardstick differs from the modified Huddart Bodenham Index, in that it is an ordered categorical classification. Furthermore, it evaluates a further two features (palatal morphology and the vertical dimension of UCL/P) compared to the modified Huddart Bodenham. The evaluation of antero-posterior and transverse dimensions are common to both and the only two features evaluated by the modified Huddart Bodenham index.

The EUROCRAN Yardstick has two elements:-

- Antero-posterior aspects (4 grades)
- Palatal aspect (3 grades)

(Appendix VII to XI)

2.9 The choice of 5-year-olds

At age 5, the outcome is based on surgical intervention alone. In contrast, measurement at 10 years of age includes other treatments such as secondary bone grafting and/or orthodontics. It is generally believed that studies of outcome for children with clefts of the lip and palate should be based on patients who are in their teens. However, this means that health care workers who look after these children would have to wait many years until the quality of treatment could be evaluated. The 'six centre intercentre study' (Shaw et al, 1992a) showed that differences could be drawn between centres at 10 years of age. But by age 10, any further treatment such as bone grafting or orthodontics dilutes the effects of surgery alone. The advantage of assessing surgical outcome on 5-year-olds is that outcome measurements is based on surgical intervention alone, lending itself to a more accurate basis for surgeons to detect differences in surgical outcome.

It is suggested that the effect of primary surgery on the maxilla and facial growth cannot be fully determined until late adolescence when the majority of facial growth has ceased. Therefore it would be better if practices such as primary surgery to lip and palate, which are detrimental to growth, could be detected even earlier. Inevitably, the ability to detect differences at an earlier stage would be highly advantageous, because it presents the surgeons with the ability to detect levels of outcome early and so forms a sound basis from which they can justify modifications of their timing or techniques. Surgery is likely to influence facial growth, dental development (Mars et al., 1992) and speech (Wyatt et al., 1996) and so an assessment at age 5, provides the best reflection of the influence of surgery on further facial growth and dental development

In order to evaluate patients surgical outcome at age 5, either prospectively or retrospectively, their records need to be readily accessible. Following Eurocleft and

CSAG recommendations, access to these records have been made feasible, as recommendations stipulate that all patients with cleft undergoing treatment have study

models for 5-years, 10-years and 18+ years of age. In Scotland the CLEFTSIS database holds study models for all patients with cleft, following these recommendations.

This give researchers the opportunity to randomly choose a sample of 5-year-old patients with UCLP. As the study models are trimmed in a standardised manner, as suggested by the Eurocleft study, a fair comparison can be drawn from this sample.

CLEFTSIS, the National Managed Clinical Network for Cleft Lip and palate Services in Scotland was formed on 1st April 2000 in response to recommendations by the CSAG report (Sandy et al., 1998). It is responsible for the organisation of a comprehensive multi-disciplinary service for the treatment of cleft lip and palate patients in Scotland. The administration is based in Perth Royal Infirmary and the members comprise the wide variety of clinicians and healthcare professionals involved in the care of cleft lip and palate patients from all Health Board areas in Scotland. All clinicians follow an agreed care pathway, a record and audit protocol, and each speciality has an agreed protocol.

An 'Electronic Patient Record' (EPR) system has been set up to overcome the problem of records being held in different sets of patients notes held in different localities by different specialties. The EPR ensures that all relevant records are taken at the correct time, so that they can be analysed and compared with standards in other countries.

At present, primary cleft surgery is undertaken in Aberdeen, Edinburgh and Glasgow. Other treatment where possible is provided near to the patients home to reduce the burden of care for these patients.

It was assumed that differences in surgical outcome could be identified in 5 year old study models, 8.5 years for a surgeon who has a caseload of 40 patients per year (Shaw et al., 1996). Early predictors of outcome provide a means to reduce the length of research studies without increasing sample size (Roberts et al., 1991).

As the recommendations were implemented in 2002, it has now been running for 6 years and in the last few years the first batch of study models will be ready for evaluation for surgical outcome. These patients will be 13.5 years old and so their records such as study models, will be readily available at age 5 & 12. It is therefore imperative that these study models are evaluated using a measuring tool, which is quick and simple to use and one which produces consistently accurate measurements. This will ensure that the measuring tool is more objective than subjective.

Summary

Future and ongoing surgical trials are likely to produce subtle, but perhaps significant results between different techniques, timing and surgeons. Therefore there is a need for a continuous rather than a categorical scale of measurement to ensure that small differences can be detected. These two indices have emerged following criticism and perceived shortfalls of the previous indices. Therefore there is now a need to compare the two emerging indices to assess which best fits the general requirements set by the World Health Organisation in 1977.

CHAPTER THREE: AIMS, OBJECTIVES & HYPOTHESIS

1. To assess and compare the reproducibility and reliability of the modified Huddart Bodenham Index and the EUROCRAN Yardstick by identifying inter- and intra - examiner variability.
2. To investigate which of the two indices is quicker to apply.
3. To investigate which of the two indices is easier to apply.

The null hypothesis is that there is no difference between the modified Huddart Bodenham Index and the EUROCRAN Yardstick.

CHAPTER FOUR: MATERIALS AND METHOD

4.1 Materials

Study models

Thirty, dental study models were selected at random from the CLEFTSiS database of 5-year old patients with non-syndromic UCLP.

Scoring forms

Scoring forms suitable for the modified Huddart Bodenham and EUROCRAN Yardstick indices were provided to all examiners, together with flowcharts and instructions for carrying out the scoring method (see Appendix I to IV and VI to XI).

4.2 Examiners

Six examiners were employed in this study, three of whom were formally calibrated in the use of a dental index and the other three with no experience of scoring indices. The examiners chosen are as follows:-

Examiner A – Professor in Restorative Dentistry

Examiner B – Specialist registrar in Orthodontics

Examiner C – Orthodontic laboratory technician

Examiner D – Laboratory technician in Restorative Dentistry

Examiner E – General Dental Practitioner

Examiner F – Nurse from Specialist Orthodontic Practice

Examiner A was calibrated for caries scoring indices and examiners B and C are calibrated for the 'Index of Orthodontic Treatment Need' (IOTN) and the 'Peer Assessment Rating' (PAR), whereas examiners D, E and F had no experience of being calibrated. Furthermore, examiner E was a General Dental Practitioner who worked solely in general practice and so had no previous nor current experience in

treating patients with cleft lip and palate. None of the above examiners, including examiner B had any previous nor present experience in treatment of cleft lip and palate.

4.3 Method

Study models

Sample selection

Thirty study models were chosen from an entire cohort of 118 5-year old patients stored on the CLEFTSiS database, using computer generated randomisation method. The sample range from 1994 to 2009.

Randomisation

The method of randomisation used was from a website as follows:-

www.mdani.demon.co.uk/para/random.htm

Open up the first link – ‘Random Number Generator & Checker’

A sentence appeared which enabled generation of 30 pseudo-random integers between 1 and 118.

At the end of this sentence, there is a drop down box and unique values were chosen. This option ensures that each integer appears ONCE only. The other options enable an integer to appear more than once.

Sample size

Statistical power calculation was not possible nor relevant for this project. Since the aim of the study was to compare two new indices, the scoring of 30 study models was based on the premise that this is a reasonable number to score in approximately one hour. The goodwill of a range of examiners was being involved in the delivery of this project and so a pragmatic approach was necessary. However, the results of this study may form the basis of a power calculation, if either of these scoring indices is to be used in future outcome studies.

Inclusion criteria

- 5-year old study models.
- Non-syndromic complete UCLP
- Good quality study models

- Randomly chosen to eliminate bias.
- Accurately trimmed in a standard manner, which follows the Eurocleft recommendation for comparative studies (Figure 10) .

Preparation for study models

All the study models were sent to Dundee Dental Hospital Laboratory where they were duplicated and prepared in a standardised method.

Preparation (Figure 10)

- Models were cast in vacuum mixed white stone
- Hand trimmed using a fine wheel to the standard height and angles
- Finished with wet and dry paper (not soaped)

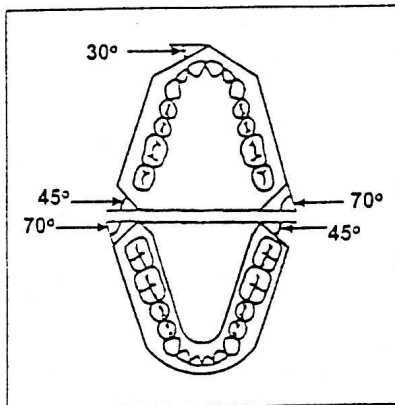


Figure A: Dental casts' base angles

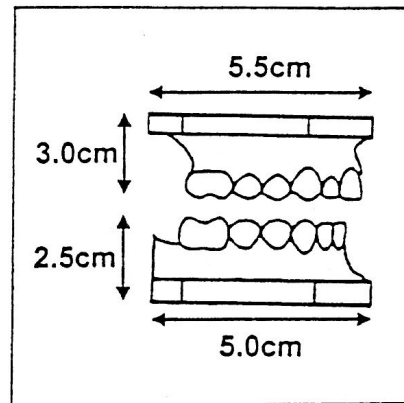


Figure B: 5-year-olds' study casts

Figure 10: Standardised preparation of study models, according to the Eurocleft recommendations

Scoring forms

4.3.1 Modified Huddart Bodenham (see appendix I to IV)

The following information was provided;

- An instruction sheet - which gives instructions on how to carry out the scoring.
- A scoring form with a list of 'exceptions to the rule' provided.
- A 'pictorial' flowchart of modified Huddart Bodenham.

A 'written' flowchart was formulated for the modified Huddart Bodenham index to make the comparison more equal between the two indices as the EUROCRAN Yardstick scoring system includes a flowchart. This would enable a fair comparative assessment of simplicity of each index.

This 'written' flowchart was piloted for consistency by assessing inter-examiner reproducibility. Four examiners were involved with the aim of showing whether there was a statistical significant difference in scores in inter-examiner reliability with the use of an extra flowchart, and thus to check if this flowchart skewed the results,

- An established 'written' flowchart designed by authors of mHB.
- Likert scale (to assess user-friendliness) (Appendix XIII)

4.3.2. EUROCRAN Yardstick (see appendix VI to XI)

The following information was provided;

- An instruction sheet - which gives instructions on how to carry out the scoring
- A list of 'Definitions of EUROCRAN Primary Dentition Yardstick'
- 'Written' flowcharts for assessing the anteroposterior, vertical and palatal forms.
- Photographs of reference models were provided to assist with the method of scoring the models.
- Scoring sheet
- Likert scale (to assess user-friendliness) (Appendix XIII)

A final checklist was enclosed at the end, to ensure that all relevant sections had been completed.

All the information from the above list was provided as published by the authors of the EUROCRAN Yardstick, and utilised as recommended. Therefore no modifications were made to this index. The instruction sheet and the Likert scale were the only two pieces of new information formulated by the authors of this project.

4.3.3 Examiners

The rationale for choosing a range of examiners was to study the reliability of applying these indices across a range of experience. It is anticipated that outcome measures, particularly in developing countries may be carried out by a range of different personnel. This will test reliability in non-expert hands which will have practical implications of scoring outcome measures in the developing countries.

The aim of selecting three calibrated and three non calibrated examiners was to investigate the simplicity of each index. This was to be assessed by recording the time taken to learn how to apply the index. This will also test the reliability of each index in non-expert hands which, will have practical implications of scoring outcome measures in the developing countries.

Instructions for carrying out the scoring method, together with the scoring forms were provided for each examiner. Thirty study models were allocated, which were to be scored in their own time. This avoids haste, which could lead to inaccuracies in the scores. Total time taken to score all the thirty study models and the time taken to learn how to apply each index was recorded. The instructions were in the same format for both indices (hard copies) and all the examiners were asked to time themselves.

Each examiner was asked to repeat scores for the same thirty study models, one month apart to enable a “washout” period. This would enable an evaluation of intra-examiner reliability.

Each examiner scored the models on four different occasions. The modified Huddart Bodenham Index was scored before the EUROCRAN Yardstick at all times to ensure

consistency in the method. Instructions for the mHB index were only given out on the day and had to be returned before repeating the scores for the Eurocran index.

This was to prevent retrospective changes being made to previous scores and an overlap of instructions, which would affect reliability of the results.

The third and fourth scores were taken at least one month after the first and second to evaluate intra-examiner reproducibility. The format for the third and fourth scoring was exactly the same as the first and second as described above.

Examiners were asked to score the models under standardised conditions, which were; to lay out the 30 models in 3 rows, with 10 in each row, in a quiet room. They did not have to score the models in one sitting as long as they timed themselves.

4.3.4 Timing

Assessors were asked to time themselves using either a stopwatch on their mobile or wrist watch. They were not asked to be specific about recording seconds.

4.3.5 Likert Scale

Once the examiners had completed the exercise they were then asked to assess the 'user friendliness' of the index by using a Likert scale (Appendix XIII).

The scale was represented by a horizontal line, divided into 10 equal sections (1cm apart). The examiners were asked to place a vertical line anywhere along this line, in order to grade their views on the use of this index (Appendix XIII).

4.4 Statistical Analysis

All statistical analysis for mHB was conducted using SPSS computer package version 15.

Statistical analysis for EUROCRAN Yardstick was conducted using Minitab.

The modified Huddart Bodenham Index

Intraclass correlation coefficient

Modified Huddart Bodenham lends itself to a continuous scale measurement and therefore a correlation coefficient can be employed to give a value of repeatability.

Repeatability is a measure of test, which assesses whether a single observer obtains the same results when s/he takes repeated measurements in identical circumstances. The Intraclass correlation coefficient (ICC) is a statistic which provides a value of agreement when continuous data are used. The ICC gives values from zero (no agreement) to 1 (perfect agreement). It is based on the means and standard deviations and it measures agreement between examiners. It is suggested that values of intraclass correlation coefficient below 0.4 indicate poor reliability, between 0.4 and 0.75 fair to good, whilst above 0.75 suggests excellent reliability (Fleiss, 1986).

Analysis/scale/reliability was chosen from the menu in SPSS and then intraclass correlation coefficient (one way random model) selected. Single measure values were noted.

Paired T Test

This statistical analysis, compares means of the two readings between the same examiner. It evaluates whether there is a statistically significant difference between the two readings. In order to calculate the confidence intervals and *p* values, the data

must first be checked for normal distribution using descriptive statistics such as boxplots. See chapter 5: Results (Figures 11-12), (Figures 19-20), & (Figure 21).

A statistical significant difference at 5% level for intra-examiner reliability can be evaluated using p values. $P > 0.05$ would indicate no statistical significant difference between the two scorings and $p < 0.05$ would indicate a difference in the two scorings.

This test is carried out using the mean differences and standard deviations of the measurements.

Bland & Altman Graphs

Bland and Altman proposed a statistical analysis which measures intra- and inter-examiner agreement (Bland, 1986). The mean differences and standard deviations are used to plot this scatter graph. The mean of the two scores is plotted against the difference between between the two scores. Mean difference and standard deviations are used to calculate the limits of agreement which are plus and minus two standard deviations from the mean difference. 95% of the differences between the two scorings should lie between the two standard deviations for good reliability. The closer the scorings lie to the line of mean difference the better the reliability of the agreement. A mean difference of zero would indicate perfect agreement.

The Bland and Altman graphs also identify random and systematic errors. The latter would be seen by a cluster of scorings which are either too high or too low. A symmetrical distribution indicates random error and therefore less chance of bias.

(Figures 13 to 18)

Student T Test

This unpaired statistical analysis, compares means of two readings between two different examiners. It also calculates a confidence interval and statistical significance at 5% level. Therefore it will be used to indicate whether there is a statistical significant difference between different examiners. Intraclass correlation coefficient will also be used to assess inter-examiner reliability.

ANOVA was not used as it assesses reliability in independent samples, where as, in this study all the samples were matched as all the examiners scored the same study models. However, there is an increased risk of getting more type I errors when multiple t-tests are used.

Wilcoxon Signed Rank Test

This is the non-parametric equivalent to the paired t-test. This was applied to assess the inter-examiner reliability for the pilot study to test the supplemental flowchart. As well as, to test for any differences, in the time taken to learn and score the indices.

EUROCRAN Yardstick

Kappa statistics

The EUROCRAN Yardstick is an ordered, categorical scale and therefore intra- and inter-examiner reliability for agreement was calculated using Cohen's unweighted kappa (κ) statistic (Altman 1991).

The unweighted (κ) statistics detects correlation between variables, i.e. it matches exact scores and if they do not match they are not correlated. So it does not take into account the degree of disagreement and therefore is a more stringent test of agreement. The weighted (κ) statistics takes into account the degree of disagreement within the results.

Furthermore, the scores for weighted kappa increase as the categories decrease. The EUROCRAN yardstick attributes 4 categories for the antero-posterior dimension and 3 categories for the palatal scores. Therefore it was decided to apply unweighted kappa statistics.

Unweighted kappa value	Strength of agreement
>0.20	Poor
0.21 to 0.40	Fair
0.41 to 0.60	Moderate
0.61 to 0.80	Good
0.80 to 1.00	Very Good

Table 2: Unweighted kappa statistics

Pilot study

A pilot study was carried out to assess the influence of an extra 'written' flowchart to supplement the existing pictorial flowchart for mHB. This was planned to enable a fair comparison with the EUROCRAN Yardstick. The number of study models used, was too small to show a normal distribution (as seen in Figure 21). Therefore a non-parametric test was carried out, the Wilcoxon Signed Rank Test.

CHAPTER FIVE: RESULTS

The reliability of the indices used for measuring surgical outcome using study models of 5 year old patients with UCLP are presented in tables and graphs below.

5.1 MODIFIED HUDDART BODENHAM

The **Intra-examiner reliability of mHB** was assessed by scoring thirty study models at two separate occasions at least one month apart. Intraclass correlation coefficient was used to assess the reliability. A score of 1 would indicate perfect agreement. The results are presented in table 3.

	Intraclass correlation coefficient
A	0.91
B	0.85
C	0.85
D	0.94
E	0.93
F	0.87

Table 3: Intra-examiner reliability for mHB assessed using intraclass correlation coefficient

The mHB index shows excellent reliability for all six examiners as with intraclass correlation coefficient ranging from (0.85 to 0.94). Half of these examiners (A, D and E) scored greater than 0.9.

Boxplots (Figures 11 & 12), show that the reliability of mHB increases with the second scores as shown by the mean (represented by a round circle with a cross) and the median. The median is represented by the middle line which indicates that 50% of the data are less than or equal to this value. Boxplots help to understand distributions. It is a graphical summary of the distribution of a sample that shows its shape, central tendency, and variability. If data is symmetrical (i.e. normally distributed), then the mean will be located half way along the box and whiskers are of similar length. The whiskers connect the largest and smallest values which are not outliers or extreme values. If the median and mean are of similar values, then it can be assumed that the data is normally distributed. The inter-quartile range box represents the middle 50% of the data, which spans from lower and upper quartiles, i.e. the 25th and 75th percentiles.

Descriptive data to show normal distribution

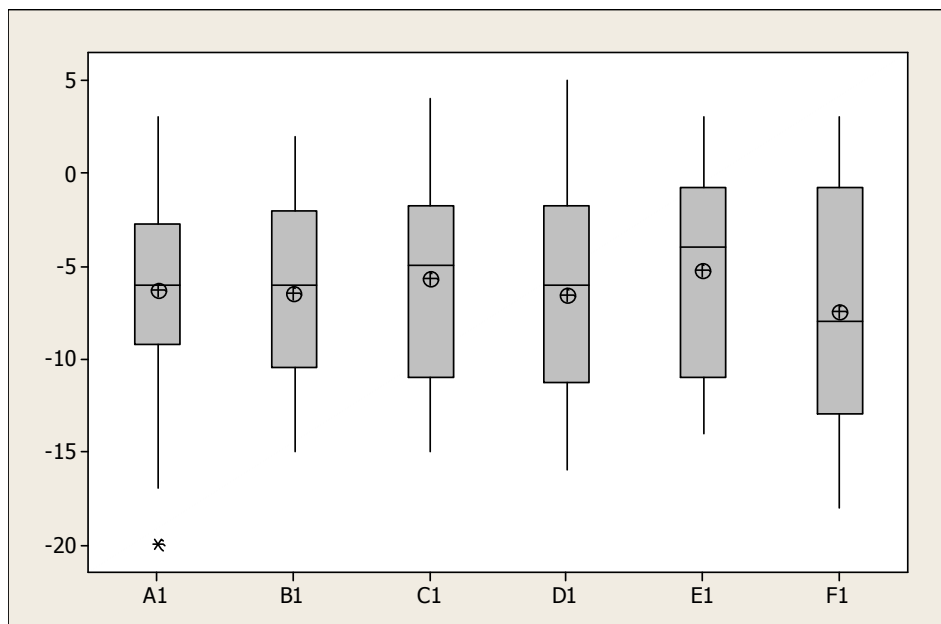


Figure 11: Boxplots to show normal distribution for first mHB scores.

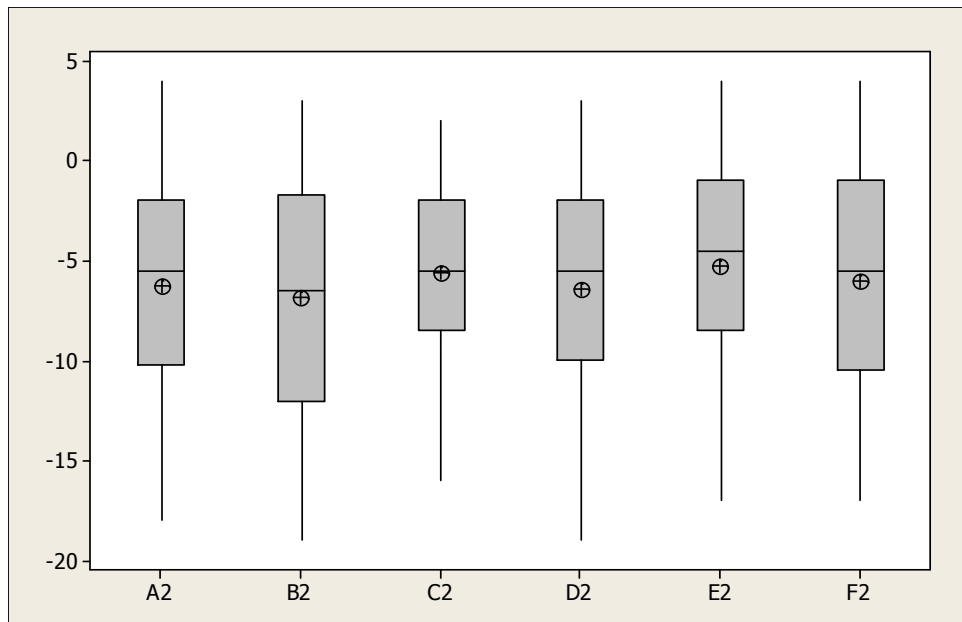


Figure 12: Boxplots to show normal distribution for second mHB scores.

Using the boxplots for the six examiners it can be shown that the data is reasonably normally distributed and that the distribution improves with the second scores. The mean and median are similar, especially with the second score as shown in figure 12.

The normal distribution for all the mHB data was further tested using the Kolmogoror-Smirnov Test, which confirms this data to be normally distributed.

Student Paired t test and 2 Sample t-test are appropriate statistical methods to use for data which are normally distributed and when the parameters such as the mean and median are approximately equal. A confidence interval and thus a p value can be calculated to show a statistically significant difference between the intra and inter – examiner reliability.

A **Paired T-Test** was applied to assess if there was a statistical significant difference between the two scores for intra-examiner reliability. No difference would favour reliability. Table 4 shows that the 95% confidence limits include zero for all the intra-examiner scores except examiner F. The p value is 0.01 and thus there is a

statistically significant difference between the two scores for examiner F and no difference for the other 5 examiners

	Mean Difference	Standard deviation	95% Confidence Interval of the difference	P value Sig. (2-tailed)
Pair 1 A1 –A2	0.07	2.22	-0.80 to 0.86	0.87
Pair 2 B1 –B2	0.37	3.03	-0.77 to 1.50	0.51
Pair 3 C1 –C2	-0.07	2.90	-1.15 to 1.02	0.90
Pair 4 D1 –D2	-0.13	1.98	-0.87 to 0.61	0.72
Pair 5 E1 –E2	-0.03	2.08	-0.81 to -0.74	0.93
Pair 6 F1 –F2	-1.47	2.90	-2.55 to -0.39	0.01

Table 4: Statistical significance for intra-examiner reliability for mHB

Bland and Altman Graphs

Bland and Altman graphs (figures 13-18) illustrate graphically agreement between intra-examiner scores for all six examiners A to F and this also identifies random and systematic errors by visual representation of score distribution.

Figure 13: Intra-examiner reliability for examiner A using mHB

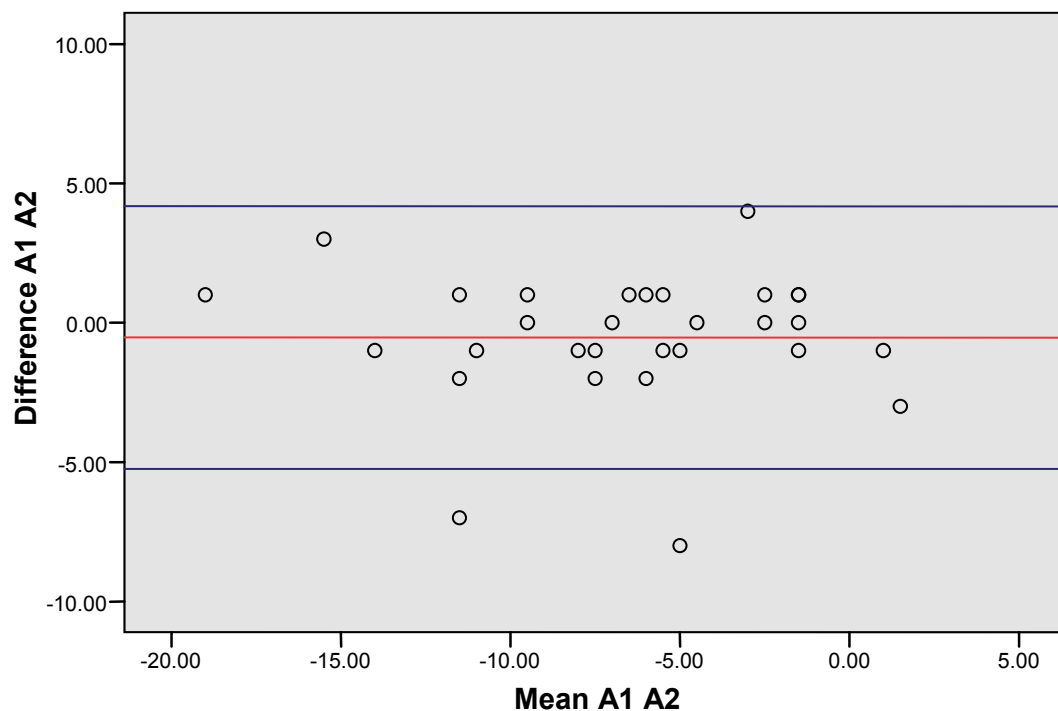


Figure 14: Intra-examiner reliability for examiner B using mHB

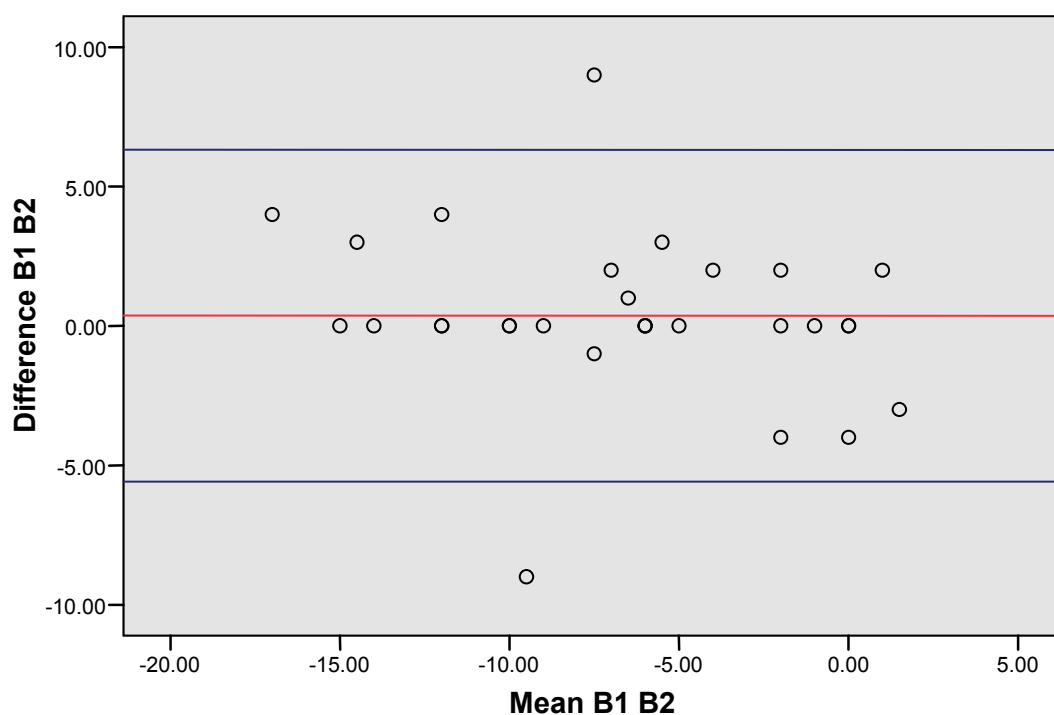
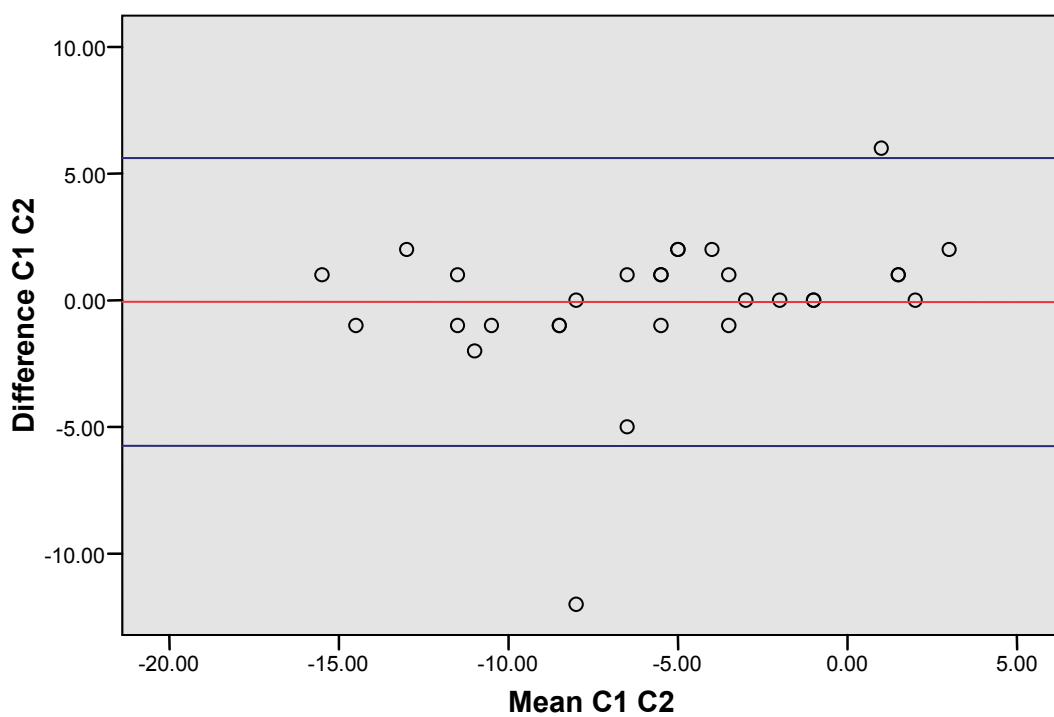
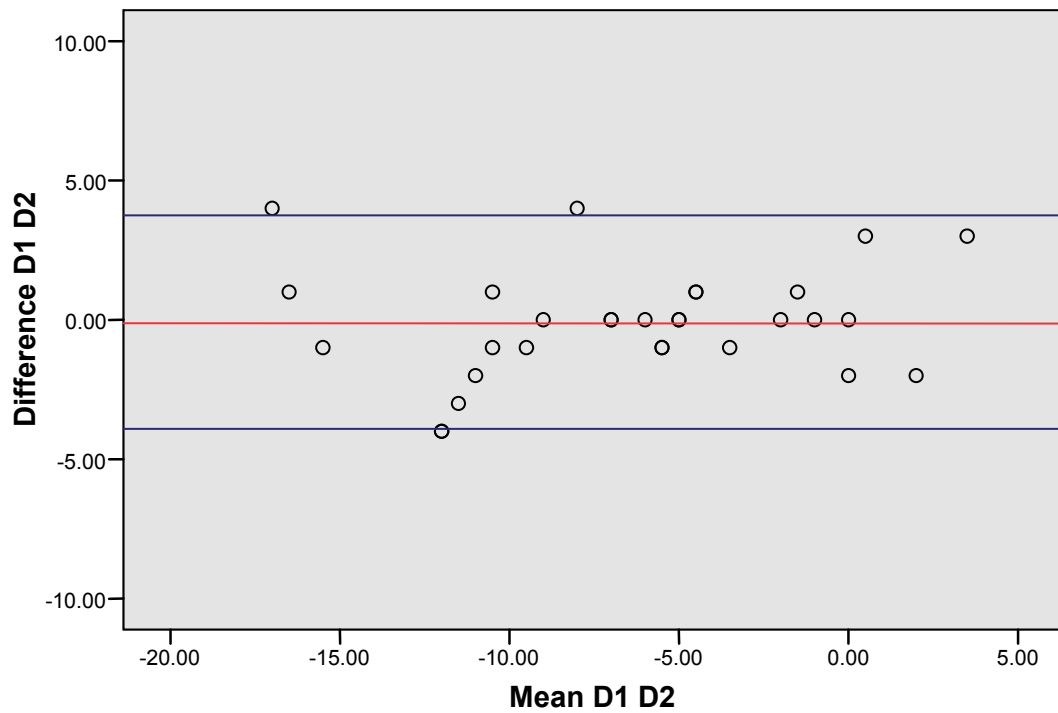


Figure 15: Intra-examiner reliability for examiner C using mHB



**Figure 16: Intra-examiner reliability for D
using mHB**



**Figure 17: Intra-examiner reliability for
examiner E using mHB**

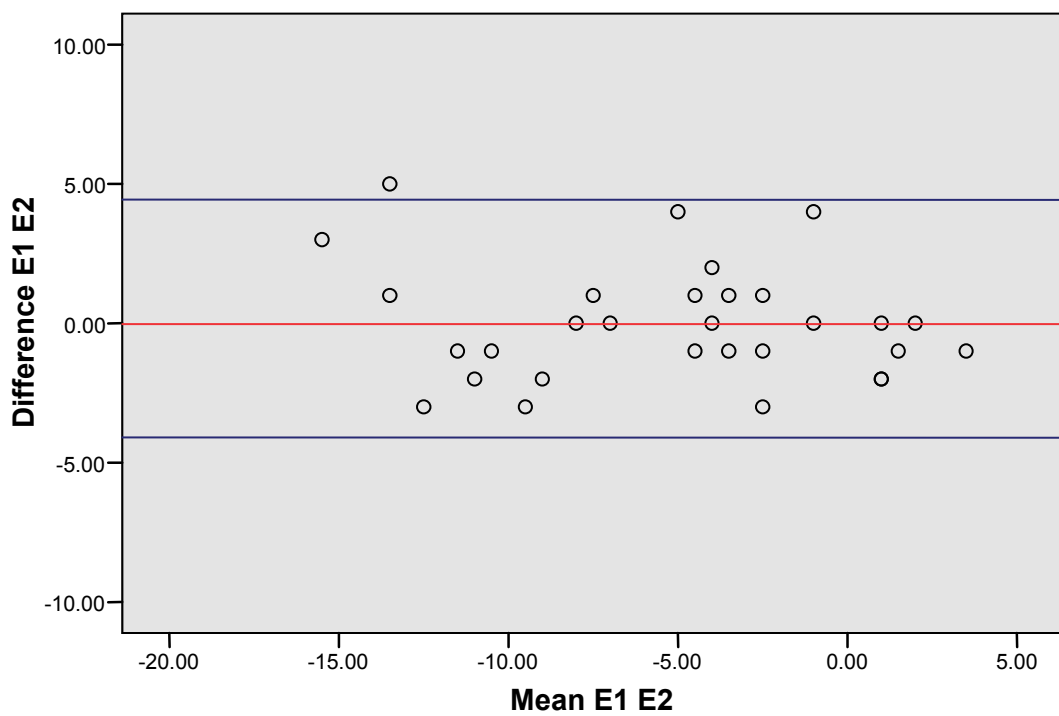
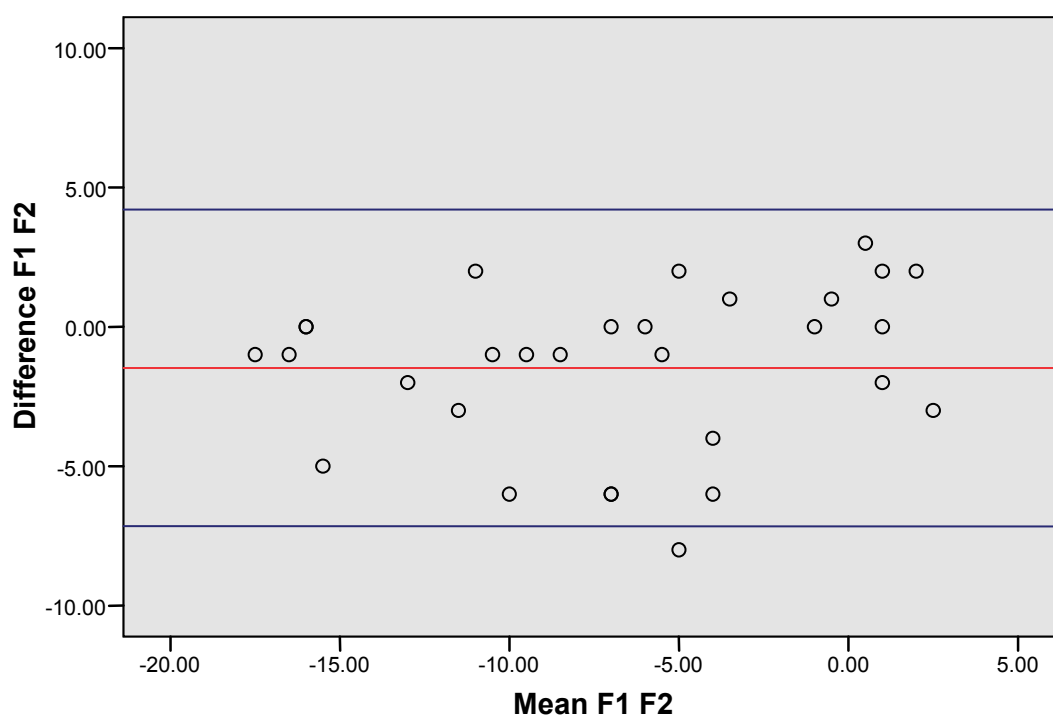


Figure 18: Intra-examiner reliability for examiner F using mHB



The graphs show that for each examiner, 95% of the results lie within ± 1.96 standard deviations which are the limits of agreement represented by the upper and lower blue lines. The central line is the mean difference between the two scores and the magnitude of the difference between the scorings is illustrated by the scatter around the line of mean difference.

Perfect agreement occurs when the mean difference is zero. This study shows that the mean difference was excellent for intra-examiner reliability between the six examiners as the mean differences are almost close to zero (Table 5). This supports the findings from intraclass correlation coefficient.

These graphs also show a good spread of data within the two standards deviation around the line of mean difference. This infers no evidence of systematic bias which would be illustrated by scores being consistently too high or low, or scores pooled in one area of the graph.

Two outliers are seen in Examiners A, B and C and one outlier in examiners E and F respectively.

The data from table 5 was used to plot the Bland and Altman graphs.

	Mean difference	Standard Deviation (SD)	Upper limit Limits of agreement	Lower limit Mean difference +/- 1.96SD
A	0.03	2.22	4.38	-4.32
B	0.37	3.03	6.32	-5.58
C	-0.07	2.90	5.61	-5.75
D	-0.13	1.98	3.75	-3.91
E	-0.03	2.08	4.43	-4.10
F	-1.47	2.90	4.21	-7.15

Table 5: Mean difference, standard deviation and limits of agreement (mean difference +/- 1.96SD)

The mean differences for 5 of the intra-examiner scores are small ranging from 0.37 to -0.53 and not clinically important. The 95% confidence limits all include zero, indicating that there is no statistically significant difference between the scores for intra-examiner reliability. However, for examiner F the mean difference is greater than the rest of the examiners and this is reflected in the wider limits of agreement for this examiner. In general all the examiners have wide limits of agreement.

Tables 6 & 7 present data for **inter-examiner reliability for mHB index**. Intraclass correlation coefficient was applied to assess the level of reliability and 2 Sample t-test to assess whether there was a statistical significant difference between inter-examiner reliability.

Intraclass correlation coefficient

Examiners	A	B	C	D	E	F
A		0.86	0.92	0.90	0.86	0.81
		0.93	0.83	0.95	0.93	0.90
B			0.81	0.83	0.81	0.77
			0.75	0.92	0.88	0.88
C				0.94	0.89	0.86
				0.80	0.90	0.84
D					0.85	0.92
					0.89	0.85
E						0.74
						0.93
F						

Table 6: Inter-examiner reliability for mHB assessed using intraclass correlation coefficient

Table 6 shows that intraclass correlation coefficient for all except three values are above 0.8 indicating that in general there is excellent inter-examiner reliability for mHB. Again, the results indicate that perhaps examiner F varies from the other examiners with scores of 0.74 between C & F and 0.77 between B & F.

p values from 2 Sample t-test

Examiners	A	B	C	D	E	F
A		0.90	0.65	0.87	0.45	0.46
		0.72	0.55	0.96	0.41	0.81
B			0.55	0.96	0.36	0.51
			0.36	0.76	0.27	0.58
C				0.54	0.78	0.25
				0.53	0.79	0.76
D					0.36	0.56
					0.40	0.79
E						0.15
						0.60
F						

Table 7: Statistical significance for inter-examiner reliability for mHB

Table 7, shows that there is no statistically significant difference between all six examiners, indicating acceptable inter-examiner reliability. However, inter-examiner

reliability between examiners E and F is almost statistically significant at p value of 0.15 at a significance level of 5%. P values in black and red scores are for first and second scores respectively.

Boxplots 19 & 20 clearly demonstrate very minimal differences in scores between Dentists and PCDs and between calibrated and non-calibrated examiners.

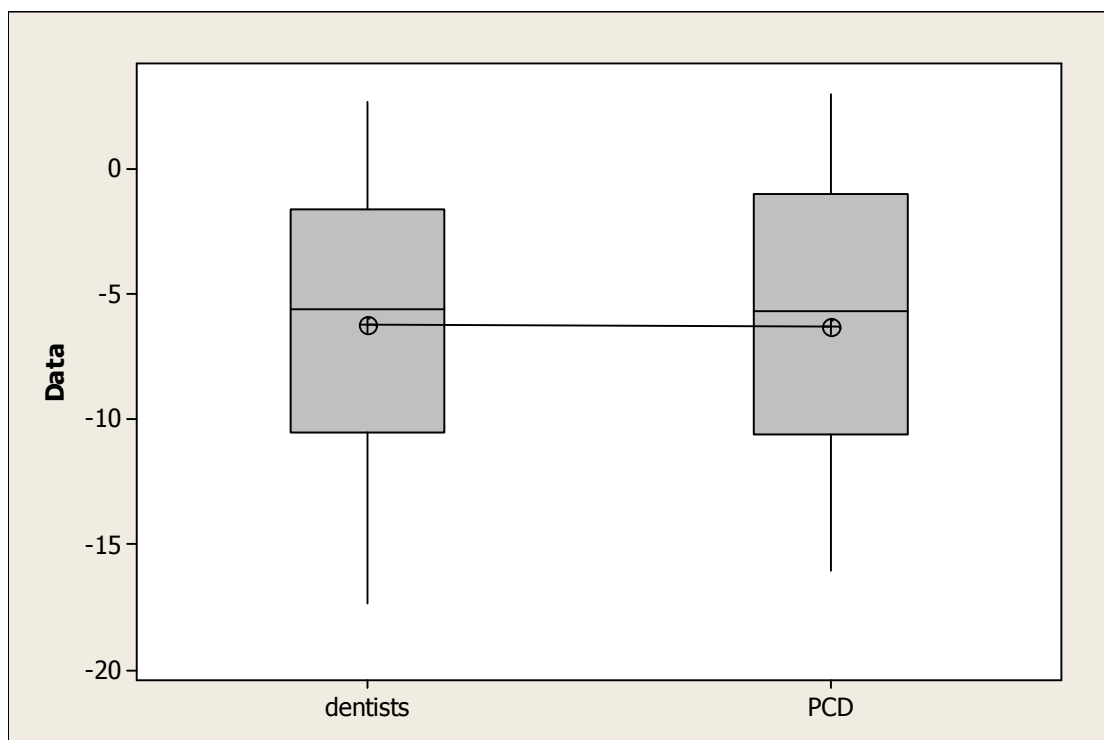


Figure 19: Boxplots to show normal distribution of data for Dentists& PCDs for mHB

	Mean	Median	SD
Dentists	-6.18	-5.59	5.19
PCDs	-6.30	-5.67	5.39

Therefore the CI for the difference = (-2.61 to 2.86) and p value =0.928

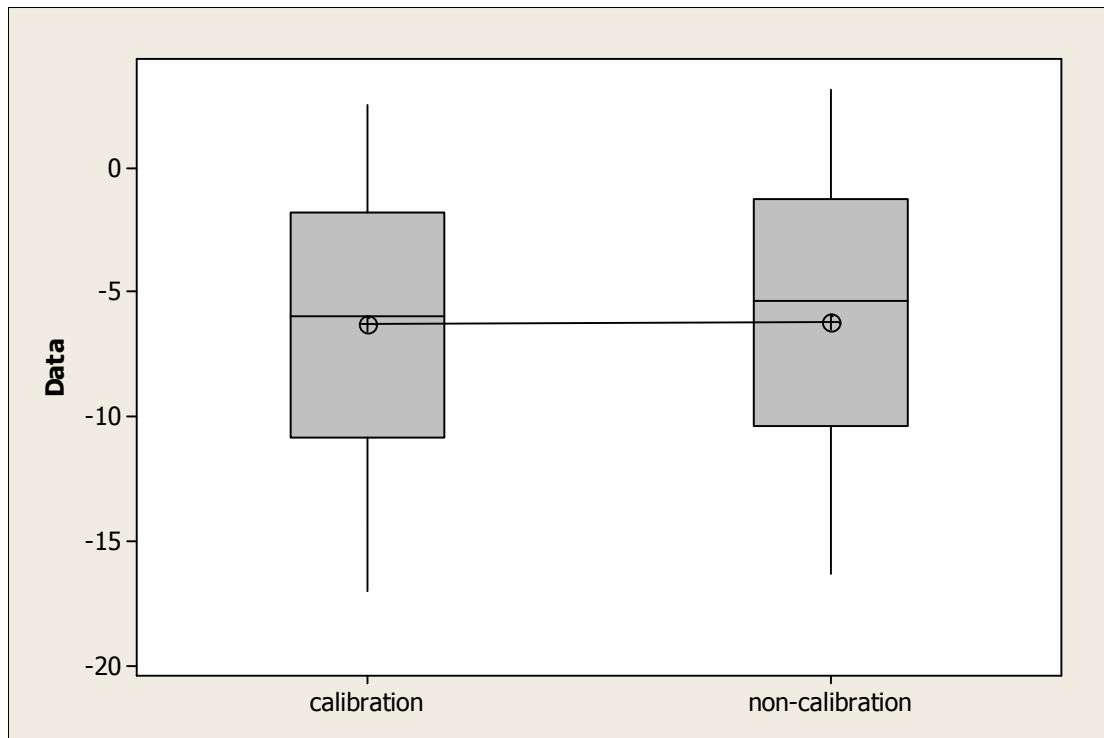


Figure 20: Boxplots to show normal distribution of data for calibrated & non-calibrated examiners for mHB

	Mean	median	SD
calibrated	-6.31	-6.00	5.12
Non-calibrated	-6.71	-5.33	5.46

Therefore the CI for the difference = (-2.87 to 2.60) and p value = 0.922

Boxplots 19 & 20 demonstrate symmetrical distribution of data about the mean (round circle with cross), which is located almost $\frac{1}{2}$ way along the box with whiskers of similar length. The median (horizontal line across the box) and mean are almost similar in value. Therefore Student T Test was carried out, which shows no statistically significant difference between the two groups. Fifty per cent of all data in all groups lie within the 25th and 75th percentiles and there are no outliers.

5.2 EUROCRAN Yardstick

Intra- and inter-examiner reliability of the EUROCRAN yardstick was assessed by scoring thirty study models in exactly the same format as the mHB. Results of this assessment are present in tables 8 & 9 and 10 and 11. These results were calculated using unweighted kappa statistics.

	A	B	C	D	E	F
A	0.60					
B		0.51				
C			0.59			
D				0.33		
E					0.60	
F						0.41

Table 8: Intra-examiner reliability of antero-posterior dimensions for EUROCRAN Yardstick

	A	B	C	D	E	F
A	0.15					
B		0.29				
C			0.50			
D				0.33		
E					0.35	
F						0.38

Table 9: Intra-examiner reliability of the palatal vault for EUROCRAN Yardstick

Unweighted k values indicate fair to moderate (0.33 to 0.60) and poor to moderate (0.15 to 0.50) reliability for the antero-posterior dimension and palatal vault.

Examiners	A	B	C	D	E	F
A		0.51	0.73	0.64	0.42	0.95
		0.48	0.46	0.34	0.44	0.41
B			0.51	0.33	0.42	0.51
			0.59	0.42	0.48	0.30
C				0.38	0.47	0.69
				0.51	0.59	0.30
D					0.38	0.59
					0.51	0.46
E						0.42
						0.53
F						

Table 10: Inter-examiner reliability of antero-posterior dimensions for EUROCRAN Yardstick

This showed a wide variation of unweighted k values which ranged from 0.33 to 0.95 for the first scores indicating poor to very good reliability and 0.30 to 0.59 for the second scores, indicating fair to moderate reliability. The best score of 0.95 was 0.41 at the second reading and similarly the next best score of 0.73 was 0.48 at the second reading. Thus the reliability is not maintained when the initial scores are good nor do they necessarily improve with repeated scoring as seen in table 5.2.4.

Examiners	A	B	C	D	E	F
A		0.07	0.23	0.20	0.21	0.15
		0.26	0.30	0.39	0.35	0.13
B			0.12	0.12	0.34	0.32
			0.38	0.27	0.21	0.31
C				0.31	0.40	0.25
				0.49	0.40	0.44
D					0.09	-0.06
					0.24	0.24
E						0.20
						0.29
F						

Table 11: Inter-examiner reliability of the palatal vault for EUROCRAN Yardstick

The unweighted k values for this aspect are the least reliable, ranging from -0.06 to 0.40 indicating poor to fair reliability for the first set of scores. However, all the scores improved with the second readings as shown in table 5.2.5. Nevertheless, the reliability only improved from fair to moderate.

5.3 Learning and Scoring

Examiners spent longer learning to use the EUROCRAN Yardstick compared to the mHB (table 12). On average the mHB took approximately half the time to learn on first exposure to the index. However, this is not statistically significant at 5% level as tested using the Wilcoxon Signed Rank test

Examiners	modified Huddart Bodenham	modified Huddart Bodenham	EUROCRAN Yardstick	EUROCRAN Yardstick
	1 st score	2 nd score	1 st score	2 nd score
A	5	5	10	5
B	1	1	9	10
C	21	9	25	10
D	17	7	5	10
E	10	16	40	15
F	5	1	8	5
Total mins	49	39	97	50

Table 12: Time taken to learn to apply each index

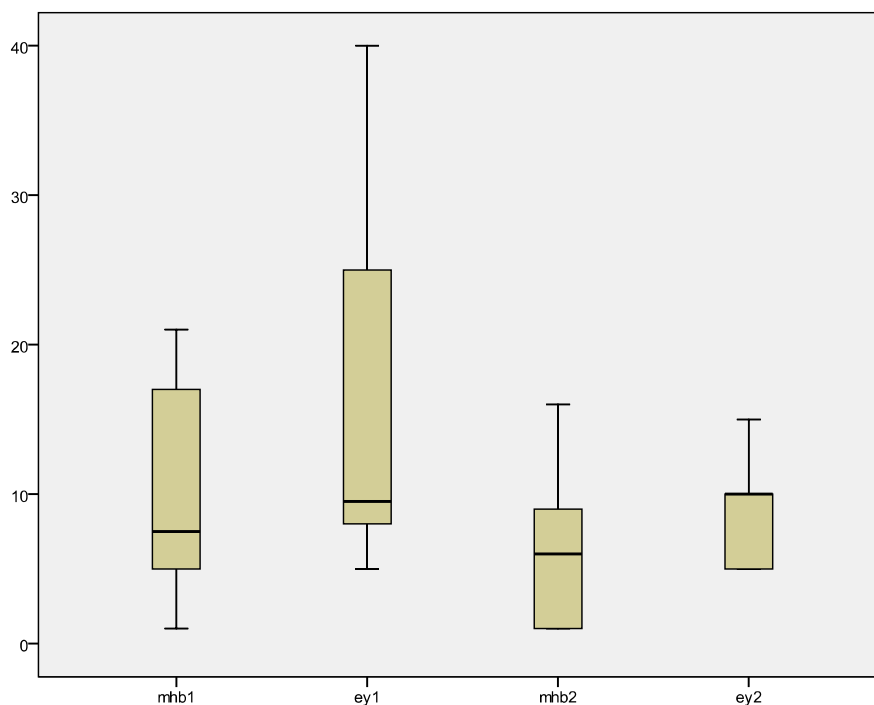


Figure A. : Boxplot to show asymmetric distribution of data from table 12

Using Wilcoxon Signed Ranks Test

mHB1 vs EY1 $p = 0.25$

mHB2 vs EY2 $p = 0.10$

Examiners spent longer to score the study models using the mHB index than EUROCRAN Yardstick (table 13) and the Wilcoxon Signed Rank Test shows that there is a statistically significant difference between the two indices.

Examiners	modified Huddart Bodenham	modified Huddart Bodenham	EUROCRAN Yardstick	EUROCRAN Yardstick
	1 st score	2 nd score	1 st score	2 nd score
A	23	36	17	18
B	31	23	23	15
C	36	28	25	16
D	50	31	27	33
E	65	45	55	25
F	65	53	35	19
Total mins	270	216	182	126

Table 13: Time taken to score each index

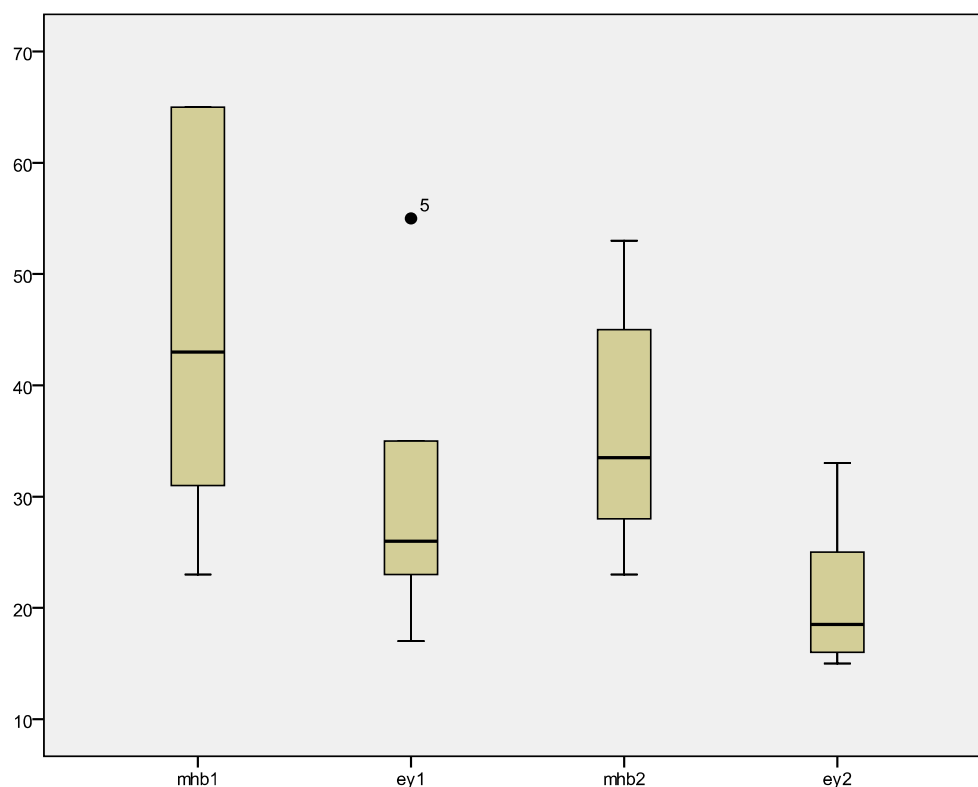


Figure B: Boxplot to show asymmetric distribution of data from table 13

Using Wilcoxon Signed Ranks Test

mHB1 vs EY1 $p = 0.03$

mHB2 vs EY2 $p = 0.05$

Table 14 shows that the total time taken for each examiner to complete the entire task. This includes both the time spent in scoring all the models in both rounds and total time spent in learning to apply the indices.

Examiners A, B and C are all calibrated and took on average 90 minutes less than the non-calibrated examiners D,E and F to complete the entire task. The dentists were on average about 4 minutes faster than PCDs.

Examiners	Total time taken for the entire task
A	119
B	113
C	140
D	180
E	271
F	191

Table 14: Total time taken for the entire exercise

5.4 User Friendliness

Table 15 shows the user friendliness of the indices used for measuring surgical outcome, measured using the Likert scale (appendix XIII)

	modified Huddart Bodenham	modified Huddart Bodenham	Eurocran	Eurocran
A	1	2	1	2
B	2	2	8	2.7
C	2.4	1	2	1.1
D	4	5	7	6
E	2.75	1.5	1.6	
F	1	1	1	1
Total	13.15	12.5	20.6	12.8 +
Average	2.2	2.1	3.4	2.6

Table 15: Likert Scale

The lower scores indicate better 'user friendliness' of the index. Overall, examiners preferred modified Huddart Bodenham compared to the EUROCRAN Yardstick.

Examiner E forgot to allocate the second score for the EUROCRAN Yardstick.

5.5 Results of the pilot study to test the supplemental flowchart

A small pilot study was carried out to assess whether the introduction of a supplemental flowchart affects inter-examiner reliability. This was assessed by asking four examiners (not the same as those chosen for the main study) to score six study models using the mHB. All the cumulative scores are presented in the table below. The boxplot (figure 21) shows asymmetric distribution of this data.

Study Models	Scorer 1	2	3	4	Average
1	-7	-11	-10	-9	-9.25
2	-8	-9	-10	-8	-8.75
3	-4	-8	-11	-6	-7.25
4	2	0	1	1	1
5	-7	-9	-6	-6	-7
6	3	2	0	4	2.25

Table 17: Data for the pilot study, using mHB to test the supplemental flowchart

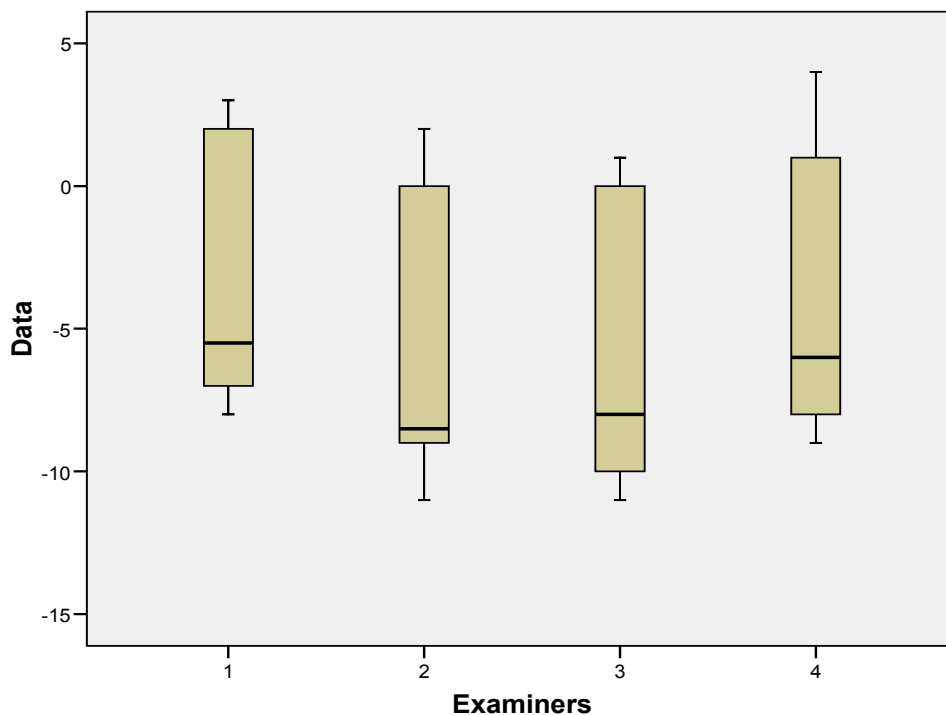


Figure 21: Boxplots to show asymmetric distribution of data for all four examiners.

<i>p</i> values		Scorer 2	Scorer 3	Scorer 4
	Scorer 1	0.03	0.06	0.33
	Scorer2		0.83	0.03
	Scorer3			0.07
	Scorer4			

Table A : To show statistical significance at 5% level for inter-examaminer reliability for the pilot study

Asymmetric distribution is a result of the small sample size and as such parametric test cannot be applied to test for statistical significance between the examiners.

Therefore a non-parametric test, Wilcoxon Signed Rank Test was applied.

This shows that a 1/3 of the reliability was statistically significant with a *p* value =0.03, concluding that the additional flowchart is not perhaps a good nor a necessary supplement to the pictorial flowchart for the mHB index.

CHAPTER SIX: DISCUSSION

6.1 Reliability

The overall results show that the modified Huddart Bodenham Index is substantially more reliable than the EUROCRAN Yardstick for intra- and inter-examiner reliability.

6.1.1 Intra-examiner reliability

Each statistical analysis, intraclass correlation coefficient, paired t-test and Bland and Altman plots demonstrate consistently high intra-examiner reliability for modified Huddart Bodenham. However, unweighted kappa scores for EUROCRAN Yardstick are comparatively low, especially for assessing the palatal aspect.

Intraclass correlation coefficient demonstrates excellent intra-examiner reliability for modified Huddart Bodenham. The paired t-test confirms this for all the examiners excluding examiner F. Bland and Altman graphs show mean differences, which are very close to zero and a symmetrical spread of data for all examiners except examiner F. Examiner F was the least experienced and was the only examiner to contact the author to clarify her understanding of the modified Huddart Bodenham Index. In order to avoid bias and maintain similarity between all the examiners, Examiner F was asked to conclude her own interpretation of the instruction sheet and score accordingly as this would influence the outcome of the study.

Bland and Altman graphs for examiners A, B, C, D and E all demonstrate random error as there is a symmetrical distribution of scores above and below the line of no difference (i.e zero). There is no evidence of systematic error in these scores.

Examiners C,D and E had the most accurate scores as they demonstrated the following:-

- random error
- no extreme outliers

- some scores on the line of no difference and majority, if not all the scores are within 2 standard deviations of it

Examiners A and B were similar to examiners C, D and E except, both had two outliers each.

Therefore, intraclass correlation coefficient is a good measure of agreement, but Bland and Altman plots demonstrate more than just agreement. It also identifies outliers, random and systematic errors and the degree of accuracy in agreement.

6.1.2 Inter-examiner reliability

Inter-examiner reliability is also consistently high for the modified Huddart Bodenham Index and low for the EUROCRAN Yardstick. The advantage of EUROCRAN Yardstick is that it can measure an extra feature (palatal vault), which strengthens the assessment of surgical outcome but the results show that the EUROCRAN Yardstick is not a reliable index for assessing this feature as the inter-examiner reliability for scoring the palatal vault was low. This may be due to the fairly vague description of the scoring of the palatal aspect and therefore, individual variation in the interpretation makes it a subjective score.

6.1.3 Calibration

Examiners A, B and C had experience in previous calibration courses, none of which involve modified Huddart Bodenham or any other cleft based index. Examiners D, E and F have no experience in calibration courses. The student t- test showed no statistically significant difference between these two groups, when applying the modified Huddart Bodenham index. Interestingly, the non-calibration examiners took more than twice as long to complete the entire task of learning and scoring the 30 study models using modified Huddart Bodenham.

6.1.4 Professionals Complimentary to Dentistry (PCDs)

Examiners C, D and F were PCDs. The student t- test showed no statistically significant difference between PCDs and dentists and boxplots show slightly better agreement between PCDs and dentists than calibrated and non-calibrated examiners for the modified Huddart Bodenham Index. Overall, PCDs took approximately 60 minutes longer to complete the entire task of learning and scoring the 30 study models using mHB.

During the assessment of inter-examiner reliability for mHB, it can be concluded that reliability is not dependent on being calibrated, and general dental practitioners, dental nurses and dental technicians are as reliable as those who are calibrated in the use of indices. However, it is necessary to acknowledge that it can take twice as long to carry out the assessment in inexperienced hands. Nevertheless, in developing countries, this finding would be very beneficial where there is a shortage of experienced personnel.

Furthermore, none of the examiners have any previous experience of treating patients with orofacial clefts. There is evidence (Atack et al., 1997b) to suggest that specialists involved in cleft lip and palate treatment can be more critical of the surgical outcome and so the results can be subject to systematic bias. However another study (Dobbyn, 2009) chose examiners who were consultant orthodontists for patients with cleft, and had excellent intra-examiner reliability with no evidence of systematic error in scoring using mHB.

6.1.5 Likert Scale

This scale was used to assess the 'user friendliness' of the two indices. The mHB was primarily rated as more user friendly than EUROCRAN Yardstick.

The comments given for EUROCRAN Yardstick are as follows:-

- *Too much to read*
- *Difficult to apply and so relying on guessing and so there is more room for error*

- *Too many aspects to remember and too many ifs and buts*
- *More difficult to learn than modified Huddart Bodenham and took longer*
- *Scoring the palatal vault was difficult*
- *Takes a long time to learn to use this index, but once it is learnt it does not take long to score the models.*
- *Scoring the palatal vault was subjective.*

The comments given for modified Huddart Bodenham are as follows:-

- *Picture flowchart was more helpful but felt that they needed both. Read the 'written flowchart' once for understanding then referred to picture chart.*
- *Read 'written flowchart' but not used it as the picture flowchart was very useful*
- *After scoring the first 10 study models, felt that it wasn't necessary to refer to the flowcharts as it was then easy to work from memory*
- *Used pictorial flowchart most of the time*
- *Having missing teeth, confused the scoring system.*

The results demonstrate that mHB is simple and straightforward to apply and does not require a calibration course. This finding is in agreement with the study by Mossey et al (2003). Although this study postulated that the study models can be accurately scored by PCDs, intra-examiner reliability for one individual (examiner F) was statistically significant at 5% level. This is as a result of a mean difference of -1.47, which was much greater than the other examiners. Intraclass correlation coefficient score for examiner F is also lower at 0.85, but not extremely different from the intraclass correlation coefficient for the other examiners. Furthermore, the Bland and Altman plot for examiner F demonstrates systematic bias, for the reason that, the scatter of scores is not symmetrical. The plot indicates that most of the scores are below zero and thus a negative mean difference. This indicates that examiner F was

consistently overestimating the severity of the surgical outcome and hence introducing systematic bias.

For mHB, the first molars are not scored even if erupted in 5-year-old study models. Therefore the maximum range of scores is -24 to +8. A mean difference of -1.47 in a range of 20 is 5%, which may be considered to be clinically significant if 5% of all the scores could be incorrectly assigned and subsequently the surgical outcome could be rated as grossly poor or excellent. This also substantiates the sensitivity of mHB in detecting small differences, which can skew the results.

Examiner F was a dental nurse with no experience of calibration therefore it could be argued that this exercise was not in their remit. However, examiner D was from the same field of experience, but was consistently high in intra-examiner reliability and displayed more accurate results with mHB than examiners A and B. Examiners A and B both have had experience in previous calibration courses and are dental professionals but no experience in treating patients with orofacial clefts.

The justification for a statistical significance for examiner F is therefore not due to their experience but due to inherent individual differences in scoring ability. Interestingly, the results show that this examiner took the least amount of time to learn to apply mHB index, yet the greatest amount of time to score the 30 study models. This could reflect on examiner F's understanding of the mHB index, which has therefore affected their scores and also the time taken to score the models. Examiners C, D and E were the most accurate and took an average of 15mins to learn to use the mHB index and approximately 30-45 minutes to score 30 study models.

Examiner B took the least time to learn to use the mHB index but had the advantage of knowing more about this index as the author of this project.

A greater understanding in the application of this index may therefore influence intra-examiner reliability and examiners should be encouraged to spend some time, approximately 15 minutes to learn to use it. The results from this project suggest that

mHB is probably not as simple to use as is accepted. Examiner F, felt that it was confusing to allocate scores when teeth were missing. Examiner F sought clarification during the exercise and she was advised to read the instructions and score according to their interpretation. This is in agreement with another study, which also identified that there may be a need for greater practice in scoring difficult areas such as canines and where there is just an alveolar ridge (Dobbyn, 2009).

An overall recommendation might therefore be to encourage participants to take time to study and understand the index prior to scoring. The study demonstrates that with repeated use the inter-examiner reliability improves and this was also demonstrated by examiner F, who was the least experienced (Table 6 & Figures 11 & 12). Therefore with the mHB, the learning curve improves with repeated use of the index.

6.1.6 Timing

According to the results it takes longer to learn to use the EUROCRAN Yardstick index compared to mHB. Conversely, it takes less time to score the models using EUROCRAN Yardstick than mHB. The increased time required to score the models using mHB is a reflection of the time taken to score each tooth with it is contra-lateral tooth in order to acquire the sum of teeth in occlusion.

6.1.7 Biases and criticism of the study design

Study Models

Thirty study models were randomly chosen, all prepared to a standardised method for a fair comparison. The plan was to start with thirty models as we were relying on the goodwill of our examiners. Table 14 indicates that most examiners took more than over two hours to complete the entire exercise. The total time spent on the entire task took between 113 to 271 minutes.

If the disparity in the reliability between the two indices should have been narrow, then it was planned to increase the sample size. However, as the differences were evidently clear, the need to increase the sample size was not necessary. Overall, the wide confidence intervals are the result of small sample size.

Scoring forms

All the forms were utilised as published and recommended by their authors. The only difference was the introduction of a narrative flowchart which was an adjunct to the pictorial flowchart for mHB. This was to enable a fair comparison between the two indices as the EUROCRAN Yardstick utilised a narrative flowchart which, summarised all the instructions on one page. There were also examples which were illustrated with photographs of the study models showing a range of severities of UCLP.

The supplemental narrative flowchart for mHB was piloted for consistency before introducing it in the main study. Four examiners were asked to assess six study models of 5-year old patients with UCLP. The introduction of this flowchart did affect inter-examiner reliability as shown by the Wilcoxon Signed Rank test. This study does not assess whether this extra modification to the mHB enhanced or deteriorated the effect of the overall reliability of this index, however there are no immediate plans to use this supplemental flowchart.

Standardised Conditions

Much effort was made to replicate standardised conditions for all six examiners. During the course of the study it was decided to allow two of the examiners to take the study models away with them to score in their own time as we realised that it was taking a lot longer than we had anticipated. This may affect a fair comparison. Even so, all the examiners were asked to lay out the study models in three rows to standardise time keeping.

6.1.8 Summary

The modified Huddart Bodenham Index is the ideal index thus far for measuring surgical outcome in 5-year-old patients with UCLP which, has been proven to be valid and reliable. It also holds other advantages which makes it distinct from other indices.

Validity

The mHB scoring system has been validated on study models, 2D photographs and clinically. It is essential that an index is not only reproducible but also valid.

Reliability

This study indicates, high reliability of the mHB index which, is reinforced by four other studies by (Mossey et al., 2003); (Gray and Mossey, 2005); (Dobbyn, 2009); (Weir, 2009).

Sensitivity

It is a sensitive index as demonstrated by this study, on grounds that, a mean difference of -1.47 in a range of 20 was judged to be clinically significant. This reinforces the findings from studies (Gray and Mossey, 2005) and (Weir, 2009).

Simplicity

This study shows that mHB is simpler to use than EUROCRAN Yardstick. However, another study (Dobbyn, 2009) showed that some training might be useful to familiarise with the mHB as this may affect reliability. Both studies show high intra-examiner reliability except that the latter study had much narrower limits of agreement. This is more likely to be due to the greater sample size rather than experience of the examiners.

Versatility

The modified Huddart Bodenham Index can be applied to any age (Mossey et al., 2003) and can successfully, be used for isolated cleft palate and BCLP with good agreement (Tothill and Mossey, 2007).

Digital recording

The use of this index was tested on digital photographs of study models and found to have good reliability (Ali et al., 2006). This would speed up the measurement and analysis of data and allow feasible inter-centre comparisons to be made.

Chairside clinical use of modified Huddart Bodenham

This index has been validated for clinical use with excellent intra- and inter-examiner reliability. This also speeds up measurement and analysis of data and eliminates the need for taking impressions, much to the patients delight.

In the future, the mHB may be used for multi-centre randomised controlled trials, which will require measurements of the antero-posterior dimensions. There is therefore good evidence to support the use of mHB as an index to measure surgical outcome for patients with UCLP. The modified Huddart Bodenham Index is also the only index to date which fulfils and has evidence to support all the requirements of an index as recommended by the 'World Health Organisation' in 1977. These are as follows:-

- 1) It should be reliable
- 2) It should be valid
- 3) It should be acceptable by the profession
- 4) It should require minimal judgement
- 5) It should lend itself to statistical analysis
- 6) It should be administratively simple

This study does suggest that the disadvantage of mHB is that it takes longer to score the study models. However, table 13 shows that there is a reduction in time at second exposure, together with an improvement in reliability at this second exposure (Table 6), thus suggesting that with repeated use, the learning curve increases as does the reliability of this index.

CHAPTER SEVEN: CONCLUSION

The results of this study do not support the null hypothesis.

The primary aim of this study was to compare the reproducibility of two indices: Modified Huddart and Bodenham Index and the EUROCRAN Yardstick.

The main conclusions from this study are as follows:

- 1) High level of reproducibility of the modified Huddart Bodenham Index compared to the EUROCRAN Yardstick
- 2) The modified Huddart Bodenham Index was deemed to be more user-friendly and therefore more acceptable than the EUROCRAN Yardstick
- 3) It is quicker to learn to apply the modified Huddart Bodenham Index.
- 4) It took longer to score the study models using the modified Huddart Bodenham Index, compared with the EUROCRAN Yardstick.

This study shows that the modified Huddart Bodenham Index is much more reproducible than the EUROCRAN Yardstick, even though it takes longer to score the models. The extra time spent is worthwhile as it produces more reliable measurements. The study does show that the time spent in scoring the study models decreases with repeated use of the modified Huddart Bodenham Index, so this should not be a significant concern.

RECOMMENDATIONS FOR FUTURE RESEARCH

A possible drawback, is that mHB does not measure or assess the palatal morphology or severity of scarring while the EUROCRAN Yardstick does. It would however be possible to record the standard 2D mirror view to obtain a photographic record of the palate both pre and post surgically.

The overall conclusion is that since previous studies have demonstrated inter-examiner reliability even from the photographs of study models (Ali et al.,2006) and it is theoretically possible to reliably score occlusion clinically using the mHB (Dobbyn, 2009), the mHB might be regarded as a viable alternative to the currently applied methods used for scoring study models i.e. the GOSLON yardstick, the 5-Year old index and the EUROCRAN Yardstick; and could potentially be used even in the absence of a study model.

This may have implications for future research and audit protocols whereby record collection could involve recording intraoral photographs of the occlusion, including a mirror (palatal) view and a direct clinical mHB score as opposed to the recording of study models. This would be applicable to both cleft palate and cleft lip/and or palate. This might have particular advantages in the developing world, where access to care is limited yet audit of quality of care is essential.

CHAPTER EIGHT: REFERENCES

- ALI, S. A., MOSSEY, P. & GILLGRASS, T. (2006) A study model based photographic method for assessment of surgical treatment outcome in unilateral cleft lip and palate patients. *European Journal of Orthodontics*, 28, 366-372.
- ALTMAN, D. G. (1991) Practical Statistics for Medical Research. *Chapman & Hall, London*.
- ASHER-MCDADE, C., BRATTSTROM, V., DAHL, E., MCWILLIAM, J., MOLSTED, K., PLINT, D. A., PRAHL-ANDERSEN, B., SEMB, G., SHAW, W. C. & THE, R. P. S. (1992) A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 4. Assessment of nasolabial appearance. *Cleft Palate-Craniofacial Journal*, 29, 409-412.
- ATAK, N., HATHORN, I., MARS, M. & SANDY, J. (1997a) Study models of 5 year old children as predictors of surgical outcome in unilateral cleft lip and palate. *European Journal of Orthodontics*, 19, 165-170.
- ATAK, N. E., HATHORN, I. S., SEMB, G., DOWELL, T. & SANDY, J. R. (1997b) A new index for assessing surgical outcome in unilateral cleft lip and palate subjects aged five: Reproducibility and validity. *Cleft Palate-Craniofacial Journal*, 34, 242-246.
- BEARN, D. R., SANDY, J. R. & SHAW, W. C. (2002) Photogrammetric assessment of the soft tissue profile in unilateral cleft lip and palate. *Cleft Palate-Craniofacial Journal*, 39, 597-603.
- BISHARA, S. E. (1973) Cephalometric evaluation of facial growth in operated and non operated individuals with isolated clefts of the palate. *Cleft Palate Journal*, 10, 239-246.
- BLAND, J. M. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1, 307-310.
- BURDI, A. R. & SILVEY, R. G. (1969) Sexual differences in closure of the human palatal shelves. *Cleft Palate Journal*, 6, 1-7.
- CHAN, K. T., HAYES, C., SHUSTERMAN, S., MULLIKEN, J. B. & WILL, L. A. (2003) The effects of active infant orthopedics on occlusal relationships in unilateral complete cleft lip and palate. *Cleft Palate-Craniofacial Journal*, 40, 511-517.
- CHEUNG, L. K. & CHUA, H. D. P. (2006) A meta-analysis of cleft maxillary osteotomy and distraction osteogenesis. *International Journal of Oral and Maxillofacial Surgery*, 35, 14-24.

- CHRISTENSEN, K., JUEL, K., HERSKIND, A. M. & MURRAY, J. C. (2004) Long term follow up study of survival associated with cleft lip and palate at birth. *British Medical Journal*, 328, 1405-1406.
- CSAG (1998) Clinical Standards Advisory Group report:cleft lip and/or palate. . *London:HMSO*.
- CUNNINGHAM, S. J. (1999) The psychology of facial appearance. *Dental update*, 26, 438-443.
- DOBBYN, L. M. (2009) Validating chairside the modified Huddart and Bodenham scoring system for cleft lip and palate M.D.Sc. *Thesis submitted to Dundee University Library*.
- EVANS (2005) Prevention and Management of Dental Decay in the pre-school child. A national clinical guideline. *Google: Sign guidelines : Guidelines: By subject: Dentistry*, 14-16.
- FERGUSON, M. W. J. (1987) Palate development: Mechanisms and malformations. *Irish Journal of Medical Science*, 156, 309-315.
- FERGUSON, M. W. J. (1988) Palate development. *Development*, 103, 41-60.
- FLEISS (1986) Design and analysis of Clinical Experiment. *New York,Wiley*.
- GORLIN, R. J., CERVENKA, J. & PRUZANSKY, S. (1971) Facial clefting and its syndromes. *Birth defects original article series*, 7, 3-49.
- GRAY, D. & MOSSEY, P. A. (2005) Evaluation of a modified Huddart/Bodenham scoring system for assessment of maxillary arch constriction in unilateral cleft lip and palate subjects. *European Journal of Orthodontics*, 27, 507-511.
- HAJEER, M. Y., AYOUB, A. F. & MILLETT, D. T. (2004) Three-dimensional assessment of facial soft-tissue asymmetry before and after orthognathic surgery. *British Journal of Oral and Maxillofacial Surgery*, 42, 396-404.
- HEIDBUCHEL, K. L. W. M. & KUIJPERS-JAGTMAN, A. M. (1997) Maxillary and mandibular dental-arch dimensions and occlusion in bilateral cleft lip and palate patients from 3 to 17 years of age. *Cleft Palate-Craniofacial Journal*, 34, 21-26.
- HOOD, C. A., HOSEY, M. T., BOCK, M., WHITE, J., RAY, A. & AYOUB, A. F. (2004) Facial Characterization of Infants with Cleft Lip and Palate Using a Three-Dimensional Capture Technique. *Cleft Palate-Craniofacial Journal*, 41, 27-35.
- HUDDART, A. G. & BODENHAM, R. S. (1972) The evaluation of arch form and occlusion in unilateral cleft palate subjects. *Cleft Palate Journal*, 9, 194-209.

- JONES, W. B. (1988) Weight gain and feeding in the neonate with cleft: A three-center study. *Cleft Palate Journal*, 25, 379-384.
- KERNAHAN (1958) A new classification for cleft lip and palate. *Plast.Reconstr.Surg*, 22, 435.
- KERNAHAN, D. A. (1971) The striped Y--a symbolic classification for cleft lip and palate. *Plastic and Reconstructive Surgery*, 47, 469-470.
- KHARBANDA, O. P., SHAW, W. C. & WORTHINGTON, H. (2002) Palate height: Another indicator of surgical outcome in unilateral cleft lip and palate? *Cleft Palate-Craniofacial Journal*, 39, 308-311.
- KRIENS (1989) LASHAL.A concise documentation system for cleft lip, alveolus and palate diagnoses. In *What is a cleft Lip and Palate? . A multidisciplinary Update*, 30-4.
- MACKAY, F., BOTTOMLEY, J., SEMB, G. & ROBERTS, C. (1994) Dentofacial form in the five-year-old child with unilateral cleft lip and palate. *Cleft Palate-Craniofacial Journal*, 31, 372-375.
- MARS, M., ASHER-MCDADE, C., BRATTSTROM, V., DAHL, E., MCWILLIAM, J., MOLSTED, K., PLINT, D. A., PRAHL-ANDERSEN, B., SEMB, G., SHAW, W. C. & THE, R. P. S. (1992) A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 3. Dental arch relationships. *Cleft Palate-Craniofacial Journal*, 29, 405-408.
- MARS, M. & HOUSTON, W. J. B. (1990) A preliminary study of facial growth and morphology in unoperated male unilateral cleft lip and palate subjects over 13 years of age. *Cleft Palate Journal*, 27, 7-10.
- MARS, M., JAMES, D. R. & LAMABADUSURIYA, S. P. (1990) The Sri Lankan Cleft Lip and Palate Project: The unoperated cleft lip and palate. *Cleft Palate Journal*, 27, 3-6.
- MARS M, P. D. (1985) Correlation of Goslon Yardstick rating and cephalometric analysis. . *Paper presented at the 5th International Congress on Cleft Palate and Related Craniofacial Anomalies, Monte Carlo*.
- MARS, M., PLINT, D. A., HOUSTON, W. J. B., BERGLAND, O. & SEMB, G. (1987) The Goslon Yardstick: A new system of assessing dental arch relationships in children with unilateral clefts of the lip and palate. *Cleft Palate Journal*, 24, 314-322.
- MATTHEWS, D., BROOMHEAD, I., GROSSMANN, W. & GOLDIN, H. (1970) Early and late bone grafting in cases of cleft lip and palate. *British Journal of Plastic Surgery*, 23, 115-129.
- MITCHELL, L. (2007) Cleft lip and palate and other craniofacial anomalies. *An introduction to orthodontics. 3rd Edition. Oxford*, Chapter 22, 244.

- MOLSTED, K., ASHER-MCDADE, C., BRATTSTROM, V., DAHL, E., MARS, M., MCWILLIAM, J., PLINT, D. A., PRAHL-ANDERSEN, B., SEMB, G., SHAW, W. C. & THE, R. P. S. (1992) A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 2. Craniofacial form and soft tissue profile. *Cleft Palate-Craniofacial Journal*, 29, 398-404.
- MORRIS, T., ROBERTS, C. & SHAW, W. C. (1994) Incisal overjet as an outcome measure in unilateral cleft lip and palate management. *Cleft Palate-Craniofacial Journal*, 31, 142-145.
- MOSSEY, P. & LITTLE, J. (2002a) Epidemiology of oral clefts: an international perspective. Wyzanski D.F. *Cleft lip and palate: from origin to treatment*, Oxford University Press, Chapter 12, 128.
- MOSSEY, P. & LITTLE, J. (2002b) Epidemiology of oral clefts: an international perspective. Wyzanski D.F. *Cleft lip and palate: from origin to treatment*, Oxford University Press, Chapter 12, 136.
- MOSSEY, P. & LITTLE, J. (2002c) Epidemiology of oral clefts: an international perspective. Wyzanski D.F. *Cleft lip and palate: from origin to treatment*, Oxford University Press, Chapter 12, 127.
- MOSSEY, P. A., CLARK, J. D. & GRAY, D. (2003) Preliminary investigation of a modified Huddart/Bodenham scoring system for assessment of maxillary arch constriction in unilateral cleft lip and palate subjects. *European Journal of Orthodontics*, 25, 251-257.
- NOLLET, P. J. P. M., KATSAROS, C., VAN'T HOF, M. A., BONGAARTS, C. A. M., SEMB, G., SHAW, W. C. & KUIJPERS-JAGTMAN, A. M. (2004) Photographs of study casts: An alternative medium for rating dental arch relationships in unilateral cleft lip and palate. *Cleft Palate-Craniofacial Journal*, 41, 646-650.
- NOVERRAZ, A. E. M., KUIJPERS-JAGTMAN, A. M., MARS, M. & VAN'T HOF, M. A. (1993) Timing of hard palate closure and dental arch relationships in unilateral cleft lip and palate patients: A mixed-longitudinal study. *Cleft Palate-Craniofacial Journal*, 30, 391-396.
- OSKOEI (2007) Web-based Learning package on The EUROCRAN Yardstick for Rating Surgical Outcome in Unilateral cleft Lip & Palate Patients. Master of Science. *Thesis submitted to University of Manchester*.
- PRAHL, C., KUIJPERS-JAGTMAN, A. M., VAN'T HOF, M. A. & PRAHL-ANDERSEN, B. (2005) Infant orthopedics in UCLP: Effect on feeding, weight, and length: A randomized clinical trial (Dutchcleft). *Cleft Palate-Craniofacial Journal*, 42, 171-177.

- PRAHL, C., KUIJPERS-JAGTMAN, A. M., VAN 'T HOF, M. A. & PRAHL-ANDERSEN, B. (2003) A randomized prospective clinical trial of the effect of infant orthopedics in unilateral cleft lip and palate: Prevention of collapse of the alveolar segments (Dutchcleft). *Cleft Palate-Craniofacial Journal*, 40, 337-342.
- PRUZANSKY, S. & ADUSS, H. (1967) Prevalence of arch collapse and malocclusion in complete unilateral cleft lip and palate. *Report of the congress. European Orthodontic Society*, 365-382.
- REID, J., KILPATRICK, N. & REILLY, S. (2006) A prospective, longitudinal study of feeding skills in a cohort of babies with cleft conditions. *Cleft Palate-Craniofacial Journal*, 43, 702-709.
- ROBERTS, C. T., SEMB, G. & SHAW, W. C. (1991) Strategies for the advancement of surgical methods in cleft lip and palate. *Cleft Palate-Craniofacial Journal*, 28, 141-149.
- ROHRICH, R. J., ROWSELL, A. R., JOHNS, D. F., DRURY, M. A., GRIEG, G., WATSON, D. J., GODFREY, A. M. & POOLE, M. D. (1996) Timing of hard palatal closure: A critical long-term analysis. *Plastic and Reconstructive Surgery*, 98, 236-246.
- RUSSELL, K. A. & MCLEOD, C. E. (2008) Canine eruption in patients with complete cleft lip and palate. *Cleft Palate-Craniofacial Journal*, 45, 73-80.
- SADLER, T. W. (2004) *Langman's medical embryology*, Lippincott Williams & Wilkins.
- SANDY, J., WILLIAMS, A., MILDINHALL, S., MURPHY, T., BEARN, D., SHAW, B., SELL, D., DEVLIN, B. & MURRAY, J. (1998) The Clinical Standards Advisory Group (CSAG) Cleft Lip and Palate Study. *British journal of orthodontics*, 25, 21-30.
- SELL, D. A. & GRUNWELL, P. (1990) Speech results following late palatal surgery in previously unoperated Sri Lankan adolescents with cleft palate. *Cleft Palate Journal*, 27, 162-168.
- SEMB, G. (1991) A study of facial growth in patients with unilateral cleft lip and palate treated by the Oslo CLP Team. *Cleft Palate Journal*, 28, 1-21.
- SEMB, G., EYRES, P., WORTHINGTON, H. & NELSON, P. (1997) Scandcleft. www.eurocran.org, Current research projects; Surgical trial.
- SHAW, W. C. (1980) Nicknames, teasing, harassment and the salience of dental features among school children. *British journal of orthodontics*, 7, 75-80.
- SHAW, W. C. (1981) The influence of children's dentofacial appearance on their social attractiveness as judged by peers and lay adults. *American Journal of Orthodontics*, 79, 399-415.

- SHAW, W. C., ASHER-MCDADE, C., BRATTSTROM, V., DAHL, E., MCWILLIAM, J., MOLSTED, K., PLINT, D. A., PRAHL-ANDERSEN, B., SEMB, G. & THE, R. P. S. (1992a) A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 1. Principles and study design. *Cleft Palate-Craniofacial Journal*, 29, 393-397.
- SHAW, W. C., DAHL, E., ASHER-MCDADE, C., BRATTSTROM, V., MARS, M., MCWILLIAM, J., MOLSTED, K., PLINT, D. A., PRAHL-ANDERSEN, B., ROBERTS, C., SEMB, G. & THE, R. P. S. (1992b) A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 5. General discussion and conclusions. *Cleft Palate-Craniofacial Journal*, 29, 413-418.
- SHAW, W. C., RICHMOND, S. & O'BRIEN, K. D. (1995) The use of occlusal indices: A European perspective. *American Journal of Orthodontics and Dentofacial Orthopedics*, 107, 1-10.
- SHAW, W. C. & SEMB, G. (1990) Current approaches to the orthodontic management of cleft lip and palate. *Journal of the Royal Society of Medicine*, 83, 30-33.
- SHAW, W. C., SEMB, G., NELSON, P., BRATTSTROM, V., MOLSTED, K. & PRAHL-ANDERSEN, B. (2000) The Eurocleft Project 1996-2000. *IOS Press Ohimha*, ISBN 1586030582.
- SHAW, W. C., SEMB, G., NELSON, P., BRATTSTRÖM, V., MØLSTED, K., PRAHL-ANDERSEN, B. & GUNDLACH, K. K. H. (2001) The Eurocleft project 1996-2000: Overview. *Journal of Cranio-Maxillofacial Surgery*, 29, 131-140.
- SHAW, W. C., WILLIAMS, A. C., SANDY, J. R. & DEVLIN, H. B. (1996) Minimum standards for the management of cleft lip and palate: Efforts to close the audit loop. *Annals of the Royal College of Surgeons of England*, 78, 110-114.
- SLADE, P., EMERSON, D. J. M. & FREEDLANDER, E. (1999) A longitudinal comparison of the psychological impact on mothers of neonatal and 3 month repair of cleft lip. *British Journal of Plastic Surgery*, 52, 1-5.
- SOAL, H. M. (2002) Classification and description of nonsyndromic clefts. *Wyzanski D.F. Cleft lip and palate: from origin to treatment*, Oxford University Press, Chapter 5, 51.
- SPRIESTERSBACH, D. C., DICKSON, D. R., FRASER, F. C., HOROWITZ, S. L., MCWILLIAMS, B. J., PARADISE, J. L. & RANDALL, P. (1973) Clinical research in cleft lip and cleft palate: the state of the art. *Cleft Palate Journal*, 10, 113-165.

- TORTORA, C., MEAZZINI, M. C., GARATTINI, G. & BRUSATI, R. (2008) Prevalence of abnormalities in dental structure, position, and eruption pattern in a population of unilateral and bilateral cleft lip and palate patients. *Cleft Palate-Craniofacial Journal*, 45, 154-162.
- TOTHILL, C. & MOSSEY, P. A. (2007) Assessment of arch constriction in patients with bilateral cleft lip and palate and isolated cleft palate: a pilot study. *European journal of orthodontics*, 29, 193-197.
- VEAU (1931) *Division Palatine*, 568.
- WEIR, J. T. (2009) Calibration of the modified Huddart and Bodenham scoring index with the GOSLON and 5-year index. M.D Sc. *Thesis submitted to Dundee University Library*.
- WOLFSBERG, E. A. (2002) The orofacial examination: normal and abnormal findings. *Wyzanski D.F. Cleft lip and palate: from origin to treatment*, Oxford University Press, Chapter 4, 43.
- WWW.WHO.INT/CLASSIFICATIONS/EN Classification of Functioning, Disability & Health.
- WYATT, R., SELL, D., RUSSELL, J., HARDING, A., HARLAND, K. & ALBERY, E. (1996) Cleft palate speech dissected: A review of current knowledge and analysis. *British Journal of Plastic Surgery*, 49, 143-149.
- ZEMANN, W., MOSSBÖCK, R., KÄRCHER, H. & KO?ELJ, V. (2007) Sagittal growth of the facial skeleton of 6-year-old children with a complete unilateral cleft of lip, alveolus and palate treated with two different protocols. *Journal of Cranio-Maxillofacial Surgery*, 35, 343-349.
- ZHENG, W., SMITH, J. D., SHI, B., LI, Y., WANG, Y., LI, S., MENG, Z. & ZHENG, Q. (2009) The natural history of audiologic and tympanometric findings in patients with an unrepaired cleft palate. *Cleft Palate-Craniofacial Journal*, 46, 24-29.